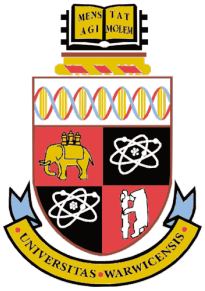
Prehospital recognition of sepsis by ambulance clinicians

(PRoSAiC)

Michael Anthony Smyth



A thesis presented for the degree of

Doctor of Philosophy

Department of Health Sciences

University of Warwick

England

31 August 2017

Dedication

To Tina, for her love and continued support, while I was shut away 'in my box', or otherwise 'on another planet'…

Acknowledgements

I came to research late in life, and several people have played a part in steering me to this point…

- Dr Joanne Fisher, who first invited me to be part of the National Ambulance Guidelines group. My journey into evidence based medicine probably started here.
- Professor Malcolm Woollard, who took one of my assignments, and turned it into real world research, generating my first two academic publications, sparking an interest in academia.
- Professor Matthew Cooke, chair of the National Ambulance Guidelines group, who invited me to join some of his research as a co-investigator, and who encouraged me to apply for my fellowship.
- My fellow research fellows at the Clinical Trials Unit, Samantha Brace-McDonnell, Dr Keith Couper, Dr Claire Hawkes and the late Dr Carl McQueen (RIP), for their support and willingness to listen to me moan about lack of progress and data related problems.
- I should not forget the unfortunate souls who have had the misfortune to share an office with me, who have suffered through my growing pains.

Lastly, my special thanks go to those who have helped bringing this thesis to fruition:

- The NIHR who funded my fellowship.
- Dr Mark Ragoo, Mr Andrew Fraser and Ms Sevilay Nizam-Basha at UHNS who extracted the hospital data.
- Mr Matthew Ward, Mrs Gill Price, Mrs Jenny Lumley-Holmes and Mr Bhu Patel at WMAS who extracted the ambulance service data.
- Mr Daniel Gallagher, who helped me better understand the syntax of R, and guided me in the use of various R packages.
- Assistant Professor Dr Peter Kimani, who helped formulate the statistical plan that became my fellowship, and for his guidance as a PhD supervisor.
- Professor Gavin Perkins, supervisor and mentor extraordinaire, who has opened doors way above my 'pay grade'. I am eternally grateful for all you have done. Did I mention I'm older than you?

Table of Contents

List of Tables

List of Figures

Declaration

This thesis is submitted to the University of Warwick, in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application, or for a degree at any other institution. The work presented (including data preparation and data analysis), was carried out by me except in the cases outlined below:

- Mrs Samantha Brace-McDonnell was second reviewer for the systematic review.
- Mrs Samantha Brace-McDonnell and Professor Gavin Perkins reviewed and recommended changes to manuscripts arising from the systematic review (Chapter three) which have been published (see below).
- Mr Daniel Gallagher provided technical guidance with respect to statistical methods and implementation of R packages.

Parts of this thesis have been published by the author:

- Smyth MA, Brace-McDonnell SJ, Perkins GD (2016) Identification of adults with sepsis in the prehospital environment: A systematic review. *BMJ Open*.
- Smyth MA, Brace-McDonnell SJ, Perkins GD (2016) Impact of prehospital care on outcomes in sepsis: A systematic review. *Western Journal of Emergency Medicine*.

The above papers are included in Appendix 1.

This dissertation contains fewer than 80000 words exclusive of appendices, bibliography, footnotes, tables and equations.

Abstract

Context: Prehospital recognition of adult patients with sepsis may inform scene management by ambulance clinicians, improve decisions concerning both appropriate hospital destination and urgency of transport, as well as facilitate early intervention before arriving at hospital.

Objective: To develop a prehospital sepsis screening tool, derived from prehospital data, for use by ambulance clinicians.

Design: The thesis comprises a systematic review of sepsis among adult patients in the prehospital environment, followed by the derivation and validation of a sepsis screening tool, utilising a retrospective data cohort comprising data from West Midlands Ambulance Service (WMAS) and the Emergency Department at University Hospital North Staffordshire (UHNS). This is followed by a comparison with alternate screening tools.

Patients: Consecutive patients transported by WMAS (n=38483) to UHNS between 01 July 2013 and 30 June 2014. Records were linked using LinkPlus® software. Successful linkage was achieved in 33289 cases (86%). Eligible patients included adult, non-trauma, non-mental health, non-cardiac arrest cases. Of 33289 linked cases, 22945 cases were eligible. The eligible cases were randomly divided into derivation (n=16063, 70%) and validation (n=6882, 30%) cohorts.

Outcome Measure: High risk of sepsis, as defined by the 2016 National Institute for Health and Care Excellence (NICE) Sepsis guideline (NG51).

Results: High risk of sepsis was present in 3.7% of both derivation (n=593) and validation (n=254) cohorts. The Screening to Enhance PrehoSpital Identification of Sepsis (SEPSIS) tool is composed of the following variables: age, respiratory rate, peripheral oxygen saturations, heart rate, systolic blood pressure, temperature and level of consciousness ($p<0.001$ for all variables). Area under the receiver operating characteristic curve was 0.87 (95%CI 0.85-0.88) for the derivation cohort, and 0.86 (95%CI 0.84-0.88) for the validation cohort. Applying a cut-off of 3 or higher, sensitivity for the SEPSIS screening tool was 0.80 (95%CI 0.74-0.84), specificity was 0.78 (95%CI 0.77-0.79), positive predictive value was 0.12 (95%CI 0.10-0.14), negative predictive value was 0.99 (95%CI 0.99-0.99), positive likelihood ratio was 3.56 (95%CI 3.30-3.85), negative likelihood ratio was 0.26 (95%CI 0.21-0.34) and the diagnostic odds ratio was 13.5 (95%CI 9.9-18.4).

Conclusion: The SEPSIS screening tool was significantly associated with high risk of sepsis status on arrival at the Emergency Department. It performs marginally better than both the UK Sepsis Trust "Red Flag" algorithm and National Early Warning Score (NEWS≥5) in an undifferentiated, adult, medical population. The SEPSIS screening tool requires external validation, in clinical practice by ambulance clinicians, in an independent population.

List of abbreviations

Abbreviation	Definition
°C	degrees Centigrade
°F	degrees Farenheit
AIC	Akaike Information Criteria
AUC	Area under the curve
AVPU	Alert, verbal, pain or unresponsive score
Bi	Regression Coefficient
BM	Blood glucose
bpm	Breaths per minute or beats per minute
CAG	Confidentiality Advisory Group
CBRT	Capillary Bed Refill Time
cc	Cubic centimeters
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CINAHL	Cummulative Index of Allied Health Libraries
CIS	Critical Illness Score
CO_2	Carbon dioxide
CRP	C-Reactive Protein
CVP	Central Venous Pressure
DARE	Database of Reviews of Evidence
DBP	Diastolic Blood Pressure
dL	Deciliter
DOR	Diagnostic Odds Ratio
ED	Emergency Department
EDIS	Emergency department Information System
EGDT	Early Goal Directed Therapy
EMS	Emergency Medical Services
EMT	Emergency Medical Technician
ePRF	electronic Patient Report Form
$EtCO_2$	End tidal carbon dioxide
ETI	Endotracheal Intubation
FCS	Fully Conditional Specification
FN	False Negative
FP	False Positive
GCS	Glasgow Coma Scale
GDP	Gavin Perkins
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
Hg	mercury
HMBG	High mobility group B1
HR	Heart Rate
HRA	Health Research Authority
ICD	International Classification of Disease
ICU	Intensive Care Unit
ID	Identification
IL	Interleukin
INR	International Normalised Ratio
IQR	Inter Quartile Range
IRAS	Integrated Researc Application System
ITU	Intensive Care Unit

IT	Information technology
IV	Intravenous
IVF	Intravenous Fluids
kg	Kilogram
kPa	Kilopascals
LMA	Laryngeal Mask Airway
MAF	Macrophage migration inhibitory factor
MAP	Mean Arterial Pressure
MAR	Missing at Random
MAS	Michael Anthony Smyth
MCAR	Missing Completely at Random
MCMC	Markov Chain Monte Carlo technique
MCP	Monocyte chemoattractant Protein
MeSH	Medical Subject Heading
MEWS	Modified Early Warning Score
MI	Multiple Imputation
MICE	Multiple Imputation by Chained Equations
min.	Minutes
MIP	Macrophage Inflammatory Protein
ml	Milliliter
MLE	Maximum Likelihood Estimation
mm	Millimeter
mmHg	Millimeters of mercury
mmol/L	Millimoles per Liter
MNAR	Missing Not at Random
NEWS	National Early Warning Score
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NLR	Negative Likelihood Ratio
NO	Nitic Oxide
non-RCT	Non-Randomised Controlled Trial
NPV	Negative Predictive Value
NRES	National Research Ethics Service
NYSIIS	New York State Identification and Intelligence System
O_2	Oxygen
OCR	Optical character recognition
OOH	Out of Hospital
OPA	Oropharyngeal airway
OR	Odds Ratio
$paCO_2$	partial pressure of Carbon Dioxide
PAI	Plasminogen-activator inhibitor
paO_2	Partial pressure of Oxygen
PLR	Positive Likelihood Ratio
POV	Privately Owned Vehicle
pPRF	Paper Patient Report Form
PPM	Predictive means matching
PPV	Positive Predictive Value
PreSep	Prehospital Early Sepsis Detection Score
PRESS	Prehospital Recognition of Severe Sepsis
PRF	Patient Report Form
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	Prospective Register of Systematic Reviews

PRR	Pattern Recognition Receptors
RCT	Randomised Controlled Trial
ROC	Receiver operating characteristic curve
RR	Respiratory Rate
RR	Relative Risk
RRV	Rapid response vehicle
SBM	Samantha Brace-McDonnell
SBP	Systolic Blood Pressure
SD	Standard Deviation
SIRS	Systemic Inflammatory Response Syndrome
SOFA	Sequential Organ Failure Score
SpO_2	Peripheral Oxygen Saturations
SSC	Survive Sepsis Campaign
$SvcO_2$	Central Venous Partial pressure of Oxygen
Temp	Temperature
TF	Tissue Factor
TLR	Toll-like Receptor
TN	True Negative
TNF	Tissue Necrosis Factor
TP	True Positive
UHNM	University Hospital North Midlands
UHNS	University Hospital North Staffordshire
UK	United Kingdom
US	United States of America
USA	United States of America
WCC	White cell count
WMAS	West Midlands Ambulance Service

Chapter 1

Introduction

1.1 Introduction

Sepsis is not a specific illness, it is a syndrome, characterised by a broadly consistent pathophysiology. This chapter presents a broad overview of sepsis. It describes the burden of sepsis and the causes of sepsis, it also provides an overview of our current understanding of sepsis pathophysiology and reports international diagnostic recommendations. Finally, it identifies the aims of this thesis.

1.2 What is sepsis?

1.2.1 Overview

The immune response to a pathogen is a complex inflammatory process that attempts to limit the spread of the pathogen and repair any tissue damage resulting from infection by the pathogen.[1] This response involves the activation of the immune system to produce both pro-inflammatory and anti-inflammatory mediators. The balance between these groups of mediators is regulated and helps to protect the host against the invading pathogens and to facilitate tissue healing. Sepsis occurs when the host response to infection becomes amplified and dysregulated.[2] Balance is lost, and infection extends beyond the infected site causing systemic functional alterations.[3]

The Society of Critical Care Medicine and the European Society of Intensive Care Medicine have coordinated an international effort to standardise the management of sepsis and define sepsis as *a life-threatening organ dysfunction caused by a dysregulated host response to infection. Organ dysfunction can be identified as an acute change in total Sequential Organ Failure Score (SOFA score) of 2 or more points consequent to the infection.*[4] In lay terms, sepsis is a life-threatening condition that arises when the body's response to an infection injures its own tissues and organs.[5]

1.2.2 Causes of sepsis

Sepsis may be the result of infection by endotoxic gram-negative bacteria, exotoxin-producing gram-positive bacteria, as well as viral, fungal, and parasitic pathogens.[6 7] A bacterial diagnosis is made in about half of sepsis cases with gram-negative bacterial infection accounting for around 60% of bacterial cases, the remainder of cases result from gram-positive bacterial infection.[2] The most common sites of infection are the lungs, the abdominal cavity, the urinary tract and primary infections of the blood stream.[8]

1.2.3 The burden of sepsis

The incidence of sepsis is difficult to quantify. Until recently, the International Classification of Diseases (ICD) coding, did not contain codes specifically for sepsis, therefore the underlying cause of sepsis was commonly recorded without noting that sepsis was present.[9] For example, if a patient became septic whilst hospitalised for pneumonia, their health care record for the episode will likely reflect the underlying reasons for hospital admission, respiratory infection, but may not include any of the ICD-10 (ICD version 10) codes that may be associated with sepsis. Furthermore, most estimates of sepsis prevalence are based upon hospital statistics and do not include patients in primary care settings, as these data are seldom collected or available.[10]

Despite the aforementioned limitations, the incidence of severe sepsis in Europe has been estimated as being between 25-38 per 100 000 population,[11] however older studies from the United States have reported rates as high as 240-300 per 100 000 population.[3] There are no conclusive data concerning incidence of sepsis in the United Kingdom (UK). The UK Sepsis Trust has identified that there are 147000 cases of sepsis per year resulting in 44000 deaths. However, a recently published report indicates that there could be as many as 260 000 cases of sepsis each year.[12] Additionally, 27% of intensive care admissions in England and Wales are for severe sepsis, and half of these patients will die.[10] Mortality data from the UK indicate that 5.1% of all deaths are definitely associated with sepsis, and up to 7.7% of all deaths may be related to sepsis.[10]

1.2.4 Pathophysiology of sepsis

The clinical manifestation of sepsis is a direct consequence of host immune cell response to pathogens. Immune cells have numerous receptors to detect the presence of a pathogen. These receptors, collectively termed pattern recognition

receptors (PRR's), include several types of toll-like receptors (TLR's) which respond to different pathogenic molecules. TLR's therefore play a central role in initiating an immune response within the host. When activated, TLR's stimulate the immune cell to produce chemical mediators which in turn will activate different components of the immune response.

Chemical mediators subsequently activate different immune cells or regulate the expression of other chemical mediators. For example, chemical mediators will activate local macrophages and neutrophils to initiate a non-specific, innate response to the infective organism, while activation of T-lymphocytes and B-lymphocytes will initiate a pathogen specific, adaptive response to the infective organism. The host response to a pathogen therefore results in a complex interplay of numerous chemical mediators which are outlined in **table 1.1**.

Table 1.1 Chemical mediators of immune response

	Mediator	**Typical effect**
Cytokines	Interleukin-1 (IL-1) Interleukin-6 (IL-6) Interleukin-12 (IL-12) Interluekin-15 (IL-15) Interleukin-18 (IL-18) Tumour necrosis factor alpha (TNF-α) Macrophage migration inhibitory factor (MAF) High mobility group B1 (HMBG1) Plasminogen-activator inhibitor type-1 (PAI-1) Interleukin-10 (IL-10)	Activate neutrophils, lymphocytes and vascular endothelium Up-regulate cellular adhesion molecules Induce prostaglandins, nitric oxide (NO) synthase and acute-phase proteins Induce fever *Note: IL-10 predominantly inhibits the above effects*
Chemokines	Interleukin-8 (IL-8) Macrophage Inflammatory Protein-1 alpha (MIP-1α) Macrophage Inflammatory Protein-1 beta (MIP-1β) Monocyte chemoattractant Protein-1 (MCP-1) Monocyte chemoattractant Protein-3 (MCP-3)	Mobilise and activate inflammatory cells, especially neutrophils Activate macrophages
Lipid mediators	Platelet activating factor (PAF) Prostaglandins Leukotrines Thromboxane Tissue factor (TF)	Activate vascular endothelium Regulate vascular tone Activate extrinsic coagulation cascade
Oxygen radicals	Superoxide radicals Hydroxyl radicals Nitric oxide (NO)	Antimicrobial properties Regulation of vascular tone

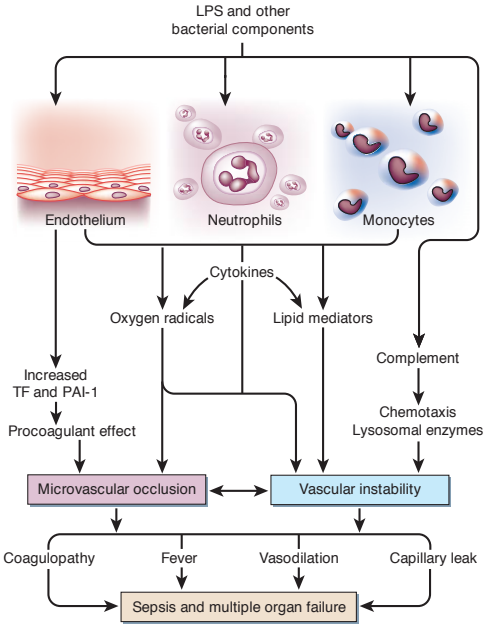
Reproduced from Cohen *et al* [2]

Activated macrophages express Interleukin-1 (IL-1), Interleukin-6 (IL-6), Tissue Necrosis Facror alpha (TNF-α), Platelet Activating Factor (PAF) as well as reactive oxygen and nitrogen species. Fibroblasts and T-lymphocytes secrete IL-6, while PAF is derived from neutrophils, macrophages, mast cells, basophils and platelets. In addition, endothelial cells also produce Tissue Factor (TF), Plasminogen Activator Inhibitor type 1 (PAI-1), nitric oxide (NO) and PAF.

IL-1 and TNF-α promote increased neutrophil adhesion and synthesis of acute phase proteins (fibrinogen, complement components, C-Reactive Protein (CRP)) in the liver. IL-6 further stimulates the synthesis of acute phase proteins (fibrinogen, complement components, CRP) by the liver as well as the production of immunoglobulins. Activation of the compliment pathway stimulates the production of complement components C3a and C5a, both of which are anaphylactoxins which leads to release of vasoactive amines (histamine, bradykinin) from mast cells and basophils.

The release of tissue factor (TF) and platelet activating factor (PAF) activates the clotting cascade, while histamine, nitric oxide and oxygen free radicals trigger capillary leakage, vasodilation and reduced myocardial contractility. These combined physiologic effects when exerted systemically lead to hypo-perfusion characteristic of sepsis (see **figure 1.1**).

Figure 1.1 Pathophysiology of sepsis



Reproduced from Cohen *et al* [2]

1.3 Diagnosis of sepsis

At present, no universally accepted diagnostic standard for sepsis exists. Three competing sets of diagnostic criteria could claim to represent the diagnostic gold standard in the United Kingdom.

The American College of Chest Physicians and the Society of Critical Care Medicine developed the original definitions of sepsis, namely the presence of infection with two or more systemic inflammatory response syndrome (SIRS) criteria.[13-15] This was the universally accepted definition of sepsis for over a decade, however research increasingly suggested this definition was overly sensitive and insufficiently specific. [16-20] Subsequently, the Society of Critical Care Medicine and the European Society of intensive Care Medicine have revised these definitions, publishing updated guidelines in April 2016, the so-called Sepsis-3 guidelines. The updated guidelines define sepsis as *a life-threatening organ dysfunction caused by a dysregulated host response to infection. Organ dysfunction can be identified as an acute change in total Sequential Organ Failure Score (SOFA score) of 2 or more points consequent to the infection.*[4]

Many health care organisations have made significant efforts to improve sepsis recognition, implementing track and trigger systems that utilise routinely measured physiologic variables, such as pulse rate and temperature. As a consequence, there has been resistance to migrate to the revised Sepsis-3 definition, which requires measurement of blood biomarkers. The National Institute for Health and Care Excellence (NICE) published their guideline [NG51] "Sepsis: recognition, diagnosis and early management" in July 2016,[21] while the UK Sepsis Trust have been advocating their "Red Flag Sepsis" guidelines since 2014.[22] Following release of the Sepsis-3 guideline [4] the Royal College of Emergency Medicine have worked with the UK Sepsis Trust to develop bundles of care for sepsis patients in the Emergency Department and suggest that patients meeting revised "Red Flag Sepsis" criteria will also meet the NICE NG51 and the Sepsis-3 criteria. [22]

1.3.1 Prehospital recognition of sepsis

Emergency Medical Services (EMS) transport a substantial number of sepsis patients.[23-25] Over half of all sepsis patients hospitalised via the emergency department (ED) will arrive via the Emergency Medical Services (EMS)[26-28] and up to 80% of severe sepsis patients admitted to intensive care via the ED will arrive by EMS.[29 30] Despite frequent exposure to patients with potentially life-threatening pathophysiology, evidence suggests prehospital recognition of sepsis by ambulance clinicians is poor.[31-36] There may be several reasons why prehospital recognition is poor, including lack of education leading to poor understanding of the condition,[17 28 37 38] ambulance clinicians may encounter sepsis cases earlier in the disease process, when the presenting condition may be less obvious,[39] lack of in-hospital diagnostic capability[23] or dependence upon SIRS criteria to formulate a diagnosis.[16 19] Several commentators have argued that a prehospital sepsis screening tool would help improve outcomes for sepsis patients.[28 40] NHS England have issued a patient safety alert, directing ambulance services to address prehospital management of sepsis.[41] There is therefore an urgent need to address prehospital identification of patients with sepsis by ambulance clinicians.

1.4 Treatment of sepsis

Management of sepsis can be challenging, but might be thought of in terms of initial basic interventions and subsequent complex support, both of which aim to improve end organ perfusion and to combat the underlying infection.[15]

Basic interventions can be delivered by non-specialist clinicians in non-specialist areas and include administration of high concentration oxygen, administration of intravascular fluids (30ml/kg) to achieve a Mean Arterial Pressure (MAP)> 65mmHg, drawing of blood to identify the causative pathogen, administration of intravenous antimicrobial therapy (preferably antimicrobials appropriate to the causative pathogen, rather than broad spectrum antibiotics), and urinary catheterisation to measure urine output. These six interventions should be achieved within the first hour and are collectively referred to as the 'sepsis six'.[42]

Once the basic interventions have been achieved, more complex support may still be required from specialist clinicians working in specialist units including obtaining central vascular access to guide further fluid resuscitation, the administration of vasopressor and inotropic drugs to support cardiac output, as

well as the administration of corticosteroids where shock persists despite fluid and vasopressor/inotropic support. In addition blood glucose levels should be monitored and actively controlled with insulin if they exceed 18mmol/L, furthermore adequate nutrition is vital to support recovery while prophylaxis to combat deep vein thrombosis and stress ulcers may be required.[43]

1.4.1 Prehospital treatment of sepsis

Several elements of the "sepsis six" can be provided by ambulance clinicians prior to arrival at hospital. Delivery of oxygen, intravenous fluids and antibiotics are already within the scope of practice for NHS paramedics.[44] However, prehospital antibiotics for sepsis are not currently indicated within the National Ambulance Guidelines.[44] Measurement of lactate is possible,[34 45-49] while studies have indicated that obtaining prehospital blood cultures is feasible.[50-52] The remaining element of the "sepsis six", measurement urine output, is not likely to be feasible in the context of an acute medical emergency. To catheterise such a patient would result in substantial, potentially life-threatening, delay to hospital.

Despite the ability to deliver most the "sepsis six", there is currently little evidence addressing management of sepsis in the prehospital environment. Several studies do show a reduction in time to key interventions and treatment tragets,[27 32 53 54] however there is little evidence of improved outcomes.[55 56] No published studies address the potential impact of prehospital antibiotic administration on further care in hospital. That is, studies published thus far have failed to include collection of prehospital blood samples prior to administration of antibiotics.

There are major implications that must be considered with respect to obtaining prehospital blood samples. First is the potential for contamination, due to poor technique and the unique challenges of the prehospital environment, resulting in misleading reporting of the causative pathogen. Second, administration of prehospital antibiotics, before obtaining blood samples, will limit the ability of hospital based laboratory services to culture the causative pathogen, and thus limit the ability of critical care services to provide targeted treatment. Responsible antibiotic stewardship must therefore be a consideration of ethical prehospital sepsis research.

Although studies conducted thus far have been small, non-randomised studies that provide weak evidence, higher quality, robust studies are underway. The PhANTASi (NCT01988428), PHRASE (ISRCTN36856873) and PITSTOP (NCT03068741) trials will hopefully provide a clearer picture of the potential impact of prehospital care on outcomes among adult patients with sepsis.

1.5 Aims and objectives

The aim of this thesis is to develop a screening tool that will improve recognition of sepsis by ambulance clinicians in the prehospital environment. The objective of developing an accurate prehospital sepsis screening tool is to facilitate delivery of appropriate interventions by ambulance clinicians, to adult patients with potentially life-threatening sepsis, before arriving at hospital. In so doing it is hoped that outcomes for this group of patients will be improved.

The thesis comprises three parts.

Part I provides the background to contextualise the study:

- Chapter one (this chapter) provides an overview of sepsis and defines the aims of the thesis.
- Chapter two describes the methodologic approach adopted to develop the SEPSIS screening tool.
- Chapter three contains a systematic review, addressing sepsis in the prehospital environment, which informed decision making in subsequent chapters.

Part II addresses development of the prehospital sepsis screening tool, derived from NHS Ambulance Service and Emergency Department data.

- Chapter four explains how the datasets used in this study were prepared prior to statistical analysis.
- Chapter five describes how the different datasets used in this study were linked together.
- Chapter six includes an overview of missing data, and details management of missing data in this study.

- Chapter seven provides a detailed description of how the SEPSIS screening tool was derived.
- Chapter eight presents the statistical validation of the SEPSIS screening tool in a clean dataset.

Part III places the SEPSIS screening tool in context

- Chapter nine addresses performance of existing screening tools in comparison to the SEPSIS screening tool.
- Chapter ten identifies the strengths and limitations of this study, and explores the research and clinical implications of the findings.

1.6 Conclusion

This chapter presented an overview of the burden, causes, diagnosis, and pathophysiology of sepsis. It has also provided an overview of the aims and objectives of this research as well also detailing the structure of the thesis. The complexity of sepsis makes diagnosis challenging; the purpose of this thesis is to simplify this challenge. Chapter two will describe the methodological approach taken in developing this thesis.

Chapter 2

Methodology

2.1 Introduction

This chapter identifies the methodology employed to develop the SEPSIS screening tool. It also provides a brief overview of how the thesis is structured and lists the permissions obtained prior to commencing any data collection or analysis.

2.2 Methodological approach

The thesis 'begins' with a systematic review. The systematic review has several purposes. To establish the incidence of sepsis within ambulance services, to draw together the available research addressing the impact of prehospital care among patients with sepsis, and importantly for future chapters, to identify potential predictors of sepsis, as well as identify existing screening tools and compare their performance.

The bulk of this thesis is dedicated to the development of the "**S**creening to **E**nhance the **P**reho**S**pital **I**dentification of **S**epsis" (SEPSIS) screening tool. Screening tools are a form of clinical prediction rule; they may also be referred to as clinical prediction models, decision rules or risk scores.[57] They estimate the probability that a specific illness is present (diagnostic model), or will respond to a form of treatment (therapeutic model), or will have a well-defined outcome (prognostic model) for an individual patient.[58]

Screening tools combine multiple predictors, drawn from patient characteristics, clinical observations, test results and disease characteristics. These are used to determine the probability that a particular outcome is present or will occur.[59 60] At heart screening tools are statistical models that inform clinical decision making, and support the delivery of evidence based medicine. Consequently, the bulk of this thesis is based entirely in the realm of quantitative statistics.

Standalone chapters are dedicated to describing the two datasets used, documenting how the datasets were prepared, and explaining the methods used to link the records in the two datasets. Management of missing data is central to quantitative research, and a chapter is devoted to multiple imputation. There

then follow three chapters addressing derivation, validation and comparison of the SEPSIS screening tool.

Regression is the most common statistical technique used to derive clinical prediction models.[59] Models developed by regression methods tend to have higher overall accuracy, that is better overall classification of all patients, but may suffer poorer sensitivity, that is, imperfect classification of atypical patients.[61]

Methodological standards for the derivation and validation of clinical decision rules have been described; originally by Wasson,[62] Feinstein,[63] Laupacis,[64] and Steill.[65] More recently Steyerberg,[59] Toll,[60] Moons,[57 66-68] Royston,[69] Bouwmeester,[70] Labarère[58] and Harrell[71] have proposed several modifications of, or updates to, these standards.

At present, there is no universally agreed methodology for the development of clinical prediction rules. However, Labarere *et al*[58] have proposed a checklist for the development and reporting of valid clinical prediction models. The checklist requires that several key elements are reported to ensure robust development of a clinical prediction model. This is the methodology that will be adhered to in this thesis (see **table 2.1**).

Table 2.1 Checklist for developing and reporting valid clinical prediction models

Item	Recommendation
Rationale	Explain the scientific background and rationale for developing a clinical prediction model
Objectives	State specific study objectives
Study design	Describe the study design (cohort, randomized controlled trial or cross-sectional), including whether patient selection was retrospective or prospective
Participants and setting	Specify the study inclusion and exclusion criteria. Report the flow of patients throughout model development. If applicable, consider use of flow diagrams complying with the STROBE statement for the derivation and validation steps and with the CONSORT extension to cluster randomized trials for impact analysis
Outcomes	Precisely define the outcome of interest in terms of timing and methods of ascertainment ("hard" outcomes are preferred)
Missing values	Report the completeness of data for each variable separately and overall for observations. Describe how missing values for predictor and outcome variables were handled in the analyses
Candidate predictors	List all candidate predictors initially considered for inclusion in the clinical prediction model (give sources of data and methods of ascertainment where relevant)

Sample size	Report rationale for sample size: The minimum number of events per candidate predictor is at least 10 for derivation studies External validation data set should contain at least 100 events and 100 non-events
Model specification	Specify statistical model building strategy, including the type of model and details of candidate predictor selection procedures
Continuous predictors	Describe how continuous predictors were handled in the analyses. If relevant, specify how thresholds were determined
Internal validation	Specify internal validation approach (split-sampling, cross-validation or bootstrapping). Bootstrapping is recommended for studies with limited effective sample size
External validation	Report all dimensions of external validation level (temporal, geographical or fully independent)
Model estimation	Derive the model on the full data set Where relevant, specify shrinkage method used in order to attenuate over-fitting Indicate whether external information was used for model updating Report the final clinical prediction model, including the regression coefficient for each predictor, along with the model intercept
Model performance	Specify how calibration and discrimination were evaluated Report apparent, internal validation and external validation performance
Model presentation	Describe model presentation for use in routine clinical practice
Model comparison	Where relevant, directly compare models developed for similar outcomes and target populations in the external validation sample
Model impact	Report effectiveness in altering intensivist practices, patient outcomes and/or costs of care. Cluster randomized trial is the preferred implementation study design
Model validity	Discuss internal (potential for over-fitting) and external (generalizability) validity and clinical usefulness

The items listed in table 2.1 are **all addressed within this thesis, however they are spread over several chapters of the thesis, and may not be addressed in the same order as listed in table 2.1.**

2.3 Study approvals

All required permissions to undertake this study were obtained via the Integrated Research Application System (IRAS project ID 152449) before any patient or clinical data was sought (see **Appendix 2**).

- A favourable ethical opinion was obtained from the National Research Ethics Service (NRES) Committee South Central - Oxford C on 02 April 2014 (REC reference: 14/SC/0163).

- Permission to process patient identifiable information without consent was obtained from the Health Research Authority (HRA) Confidentiality Advisory Group (CAG) under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 on 19 May 2014 (CAG reference: CAG4-03(PR2)2014).
- A substantial amendment, submitted to reflect the HRA CAG approval to process patient identifiable information without consent, was approved by the NRES Committee South Central - Oxford C on 06 August 2014 (Amendment number: PRoSAiC Amendment 1).
- Site agreements were confirmed with West Midlands Ambulance Service NHS Foundation Trust on 05 January 2015 and with University Hospital North Midlands NHS Trust on 19 January 2015 respectively.
- Formal approval to commence the study was provided by the National Institute for Health Research (NIHR) Clinical Research Network West Midlands on 09 April 2015.
- The study was sponsored by the University of Warwick.
- The study was funded by a Clinical Doctoral Research Fellowship (CDRF2012-05-58) awarded by the National Institute for Health Research (NIHR).

2.4 Conclusion

This chapter has provided an overview of the methodology used to develop the SEPSIS screening tool. It also provided a list of the permissions obtained prior to commencing any data collection or analysis. The next chapter documents the systematic review supporting the study.

Chapter 3

Systematic review of sepsis in adult patients in the prehospital environment

3.1 Introduction

This chapter reports a systematic review of sepsis in the prehospital environment. It addresses themes of importance to this thesis, namely to report the incidence of sepsis in the prehospital environment, to identify predictors for sepsis in the prehospital environment and to determine if screening tools improve recognition of patients with sepsis in the prehospital environment.

The systematic review will also help to establish if early identification of sepsis in the prehospital environment has any impact on outcomes among adult patients with sepsis and finally, to determine if care delivered by ambulance clinicians in the prehospital environment leads to better outcomes among adult patients with sepsis. Elements of this systematic review have been published and are included in **Appendix 1**.

3.2 Background

Severe sepsis presents a major healthcare challenge. It results in significant morbidity and mortality and carries a huge financial burden. Twenty seven percent of admissions to intensive care in England and Wales are due to sepsis with almost half of these cases being fatal, and patients admitted to intensive care due to sepsis having longer hospital stays than those admitted for other causes.[72] Accurate mortality statistics are difficult to ascertain as most published data are based upon in-hospital data.[10] In the United Kingdom (UK) the UK Sepsis Trust have identified that there are 147000 cases of sepsis each year, resulting in 44000 deaths.[5] However, a recently published report indicates that there could be as many as 260 000 cases of sepsis each year.[12] Mortality statistics for England and Wales would seem to support these estimates. McPherson *et al*[10] reported that, in 2010, 5.1% of all deaths were definitely associated with sepsis, while up to 7.7% may have been associated with sepsis.

With respect to the financial burden of sepsis, Burchardi *et al*[11] report that the cost of sepsis was estimated to fall between €23000 and €29000 per case, in European countries. Daniels *et al*[42] estimate that the cost of sepsis to the National Health Service (NHS) exceeds £2 billion each year.

Recent years have seen greater focus on improving outcomes from sepsis. Hospital data suggests that early recognition and initiation of treatment leads to improved patient outcomes. Gaieski *et al*[73] reported that the time to administration of appropriate antimicrobials was the primary determinant of mortality in patients with severe sepsis and septic shock, noting a significant association between mortality and administration of antimicrobials within the first hour of ED arrival (mortality 19.5 vs. 33.2%; odds ratio (OR) 0.30 95% CI 0.11–0.83, p=0.02). Similarly, Daniels *et al*[42] demonstrated that the delivery of early interventions was associated with reduced mortality - for patients receiving the sepsis six bundle (six predefined interventions to be implemented within the first hour) mortality was 20.0% compared with 44.1% for those patients who did not receive the sepsis six bundle (p<0.001). Where resuscitation is delayed, evidence suggests outcomes become worse. Kumar *et al*[74] reported that for each hour delay to administration of antibiotic therapy, mortality increased by 7.6% for sepsis patients with hypotension. Despite these data indicating an association between time to antimicrobial therapy and outcomes, some uncertainty regarding the optimal timeframe for delivering particular treatments and interventions remains.[75-78] There is, however, consensus that delays to recognition of sepsis and subsequent resuscitation will likely be detrimental to patient outcomes.[79-81]

The burden of sepsis upon UK ambulance services is not well understood. Guerra *et al*[32] identified that 6.9% of emergency medical services (EMS) transports in the United States are for patients with infection. Furthermore, it is estimated that 8-10% of EMS patients who have infection will be diagnosed with sepsis[24 32] Following a 10-year observational study, Seymour *et al*[25] reported the incidence of severe sepsis in a North American ambulance service to be 3.3 per hundred ambulance transports. They also noted that the incidence of severe sepsis among EMS patients was increasing at 11.8% per year, from around 1.4% in 2000 to 4.2% in 2009. Over half of all sepsis patients hospitalised via the emergency department (ED) will arrive by EMS[26-28] and up to 80% of severe sepsis patients admitted to intensive care via the ED will arrive by EMS.[29 30]

Although few comparable UK ambulance service data exist, a recent Scottish study involving 20 out of 25 mainland district general and teaching hospitals revealed that for patients who had severe sepsis or septic shock, the proportion arriving at hospital by ambulance was as high as 88.1%[30] These data suggest that ambulance services frequently provide treatment and transport of sepsis patients, and furthermore that patients with life threatening sepsis are more likely to arrive via ambulance. Consequently, there is an opportunity for ambulance clinicians to aid the early recognition of critically ill sepsis patients and potentially reduce time to intervention for this population in the same manner as they do with other time critical, life threatening conditions such as acute myocardial infarction, stroke and major trauma.[82-91]

In September 2014 NHS England issued a Stage 2 Patient Safety Alert[41] outlining how Ambulance Services could contribute to improved outcomes for sepsis patients. Key to this strategy is recognition of the septic patient by the ambulance clinician. However, it has not yet been established if ambulance clinicians are able to identify sepsis patients when access to advanced diagnostics such as laboratory analysis services are limited. Little work has been undertaken to improve sepsis recognition and management by ambulance clinicians. The purpose of this systematic review is to collate the existing evidence pertaining to prehospital recognition of sepsis and to determine if ambulance services might be able to improve outcomes for sepsis patients.

3.3 Methods

3.3.1 Review questions

The purpose of the systematic review is to collate the available evidence regarding identification, management and outcomes among adult patients with sepsis, when cared for by ambulance clinicians in the prehospital environment. The specific objectives are:

- Establish the incidence of sepsis in the prehospital environment.
- Identify known predictors for sepsis in the prehospital environment.
- Determine if ambulance clinicians can identify patients with sepsis in the prehospital environment.
- Determine if prehospital sepsis recognition by ambulance clinicians leads to improved processes of care among adult patients with sepsis.
- Determine if prehospital care delivered by ambulance clinician improves outcomes among adult patients with sepsis.

3.3.2 Review methods

The protocol for this systematic review is registered with the International Prospective Register of Systematic Reviews (PROSPERO) (CRD42014007654). It has been conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[92]

3.3.3 Identification of studies

The electronic databases MEDLINE, EMBASE, CINAHL and the Cochrane Library (includes Database of Systematic Reviews and the Database of Abstracts of Reviews of Evidence (DARE)) were systematically searched. Search strategies were developed in consultation with a medical information specialist using a combination of Medical Subject Headings (MeSH) (or equivalent) and text terms relating to both sepsis and ambulance clinicians as outlined below:

```
(Sepsis OR septic OR septic?emia OR systemic adj
inflammatory adj response adj syndrome OR SIRS OR
septic adj shock OR hypotension adj induced adj
hypoperfusion OR cryptic adj shock OR bacterial adj
infection)
AND
(emergency adj medical adj service OR EMS OR HEMS
OR emergency adj medical adj technician OR EMT OR
paramedic OR pre-hospital OR prehospital OR pre adj
hospital OR out-of-hospital OR out adj of adj
hospital OR OOH OR Ambulance).
```

The above MEDLINE search strategy was adapted for each of the databases searched, for example by modifying subject headings, but retaining the same text terms. Searches were not restricted by year of publication, study design or language of publication. Primary searches, described in detail later, were undertaken in July 2014 and were repeated in June 2015. Citations were exported into the bibliographic software EndNote® version X7 (Thompson Scientific, Carlsbad, CA), duplicate citations were removed manually within EndNote®. The citations were exported from EndNote® into a Microsoft Excel 2013® spreadsheet (Microsoft Corporation, Redmond, WA) and distributed to two reviewers, Michael Smyth (MAS) and Samantha Brace-McDonnell (SBM), for screening.

3.3.4 Selection of studies for inclusion

A priori inclusion and exclusion criteria were used to screen the retrieved citations:

3.3.4.1 Inclusion criteria

- Language: no restrictions were placed, however studies without an English full text document were not included in the systematic review.
- Publication type: original research published in peer reviewed journals
- Study Design: systematic reviews, meta-analyses, randomised controlled trials (RCT's), case-control studies, cohort studies, cross-sectional studies, retrospective analyses and conference proceedings.
- Study Population: adult participants. Populations could comprise a mix of adult and child participants if results are reported separately.
- Case Definition: no restrictions as to severity of sepsis.

3.3.4.2 Exclusion criteria

- Publication type: narrative reviews, letters, editorials, commentaries, books and book chapters, lectures and addresses, and consensus statements.
- Study Design: case reports, qualitative studies, non-systematic reviews, studies that fail to report their methods.
- Study Population: mixed adult and child population without distinct reporting, child population, animals

3.3.5 Critical appraisal

All eligible papers were independently appraised by two reviewers (MAS and SBM). To control for subjective appraisal, a standardised approach to quality assessment was adopted. Each study was assessed for risk of bias, any inconsistency of results, the indirectness of the evidence, any imprecision in the results, and any other factors such as publication bias, or dose effect that might influence the quality assessment as recommended in the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach for rating the quality of evidence.[93]

- **Risk of bias** limitations relate to issues with study design and how these limitations might affect confidence in the estimate of effect reported in the study.[94] Within the hierarchy of evidence, well conducted randomised controlled trials limit the potential for bias and

may therefore be at low risk for bias. However, factors such as lack of allocation concealment, lack of blinding, a large loss to follow-up, trials stopped early for benefit or selective reporting of outcomes will all introduce the potential for bias and thus reduce our confidence of the reported estimate of effect. Non-randomised studies are generally limited with respect to what can be inferred from results, and how generalisable the findings are to the wider population of interest. Non-randomised studies are usually therefore at significantly higher risk for bias.

- **Indirectness** of results is closely related to the populations studied.[95] The populations of several included studies contain patients who do not have sepsis. It is therefore appropriate to be very cautious in generalising the findings of these studies to the population of interest (adult sepsis patients).
- **Inconsistency** of results refers to widely differing estimates of effect between studies (heterogeneity or variability in results).[96] Where such differences exist, it is important to determine if there is a justification for the observed differences. Differing estimates could be attributable to severity of disease, timing of intervention, dosage of drug used or other many other factors. When heterogeneity exists, and influences the interpretation of results, study authors have a responsibility to identify a plausible explanation, or the risk of bias increases and the quality of evidence decreases. If estimates of effect vary considerably, without underlying reasons for the observed difference, extreme caution must be exercised, when generalising to the wider population.
- **Results** can be considered to be imprecise when studies include few participants and few events and thus have wide confidence intervals.[97] Low numbers of participants or events limit the accuracy of the calculated estimate of effect.
- **Publication bias** refers to the selective reporting of outcomes or failure to fully report estimates of effect e.g. by reporting a point estimate without the confidence interval.[98] The heavy involvement of commercial sponsors in trials also raises questions of whether unpublished data, suggesting no benefit, exists.[99]

The GRADE approach stratifies quality of evidence into four categories; 'high quality', 'moderate quality', 'low quality' and 'very low quality'.[93] Randomised

controlled trials are initially rated as 'high quality', while non-randomised trials studies are initially rated as 'low quality'. Following critical appraisal each reviewer independently adjusted the quality assessment upward or downward depending upon the risk of bias, inconsistency of results, indirectness of the evidence, imprecision in the results, or other pertinent factors.[94 100]

3.4 Results

3.4.1 Systematic literature search

The systematic search of the literature identified 4366 citations. After duplicates were removed, 2958 unique citations remained. Applying *a priori* selection criteria, two reviewers (MAS and SBM) independently rated each citation title as 'include', or 'exclude'. Citations rated as 'include' by at least one reviewer were considered potentially relevant, while citations rated as 'exclude' by both reviewers were considered irrelevant. Seventy-eight citations were identified as potentially relevant, while 2880 citations were discarded as irrelevant. One additional study,[33] a manuscript pending publication, was identified by contacting Ambulance Service Medical Directors and subject experts. Inter-rater agreement for primary citation screening, to include or exclude studies, calculated using Cohen's kappa statistic was 0.87 (95% CI 0.81 to 0.92; p<0.001).

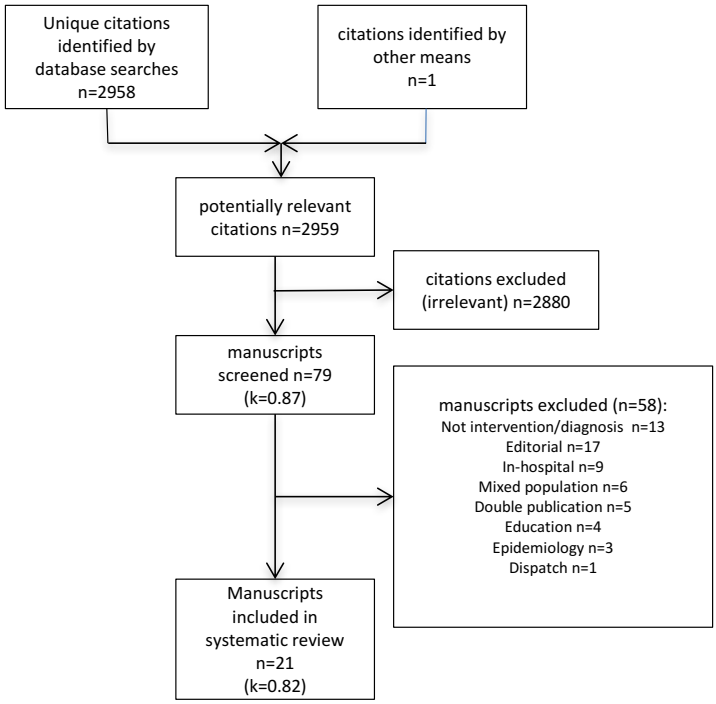
During the second stage of screening, the same two reviewers (MAS and SBM) independently reviewed the full manuscript for each of the 79 potentially relevant studies. Each study was rated as 'include', 'maybe' or 'exclude'. If both reviewers rated a manuscript as 'include' it was automatically considered for full text review. Similarly, if both reviewers rated the manuscript as 'exclude' it was automatically discarded. If there was disagreement between reviewers, for example, one rated the manuscript as 'maybe' while the other rated the same manuscript as 'exclude', the reviewers met to discuss their reasons for rating the manuscript as they did, and attempt to seek a consensus. If the two reviewers were not able to achieve consensus, a third independent reviewer (Gavin Perkins (GDP)) was available to adjudicate.

Seventy-nine manuscripts were independently reviewed. There were no instances where consensus could not be reached by the two reviewers. Fifty-eight manuscripts were discarded following the second stage screening, with 21 manuscripts included in the systematic review. Inter-rater agreement for

second-stage screening, calculated using Cohen's kappa statistic, was 0.82 (95% CI 0.68 to 0.97).

Reference lists of included manuscripts were examined to identify any missed studies. No additional items were identified by this approach (see **figure 3.1**).

Figure 3.1 PRISMA flow chart



3.4.2 Characteristics of included studies

Most studies identified involve small numbers of participants, without control and intervention cohorts. Only one randomised controlled trial was identified, with the remainder being non-randomised studies. There was considerable variation in the diagnostic standards utilised, reported outcomes and methodological approach adopted across the studies. Because of these differences, meta-analysis was not appropriate and a narrative approach to data synthesis was adopted.

Key characteristics of the 21 included studies are presented in **table 3.1**. A clear majority (71%) originate from the United States, all studies were published in English, six studies were published in abstract form only. There was one randomised controlled trial (published in abstract only); the remaining studies were all non-randomised studies.

Table 3.1 Characteristics of included studies

Characteristic	Details
Median year of publication [range]	2012 [2010–2015]
Country of publication [n (%)]	
• *United States*	15 (71)
• *Germany*	2 (9)
• *United Kingdom*	1 (5)
• *Sweden*	1 (5)
• *Australia*	1 (5)
• *Canada*	1 (5)
Language [n (%)]	
• *English*	21 (100)
Study design [n (%)]	
• *Randomised Controlled Trial*	1 (5)
• *Non-randomised Study*	20 (95)
Publication type [n (%)]	
• *Full publication*	15 (71)
• *Abstract publication*	6 (29)

3.4.3 Assessment for risk of bias

For the randomised controlled trials, risk of bias was assessed across the following domains: lack of allocation concealment, lack of blinding, incomplete accounting of patients and outcome events, selective outcome reporting bias and other limitations such as stopping a trial early for benefit. Bias within non-randomised studies was assessed across the domains of failure to develop and apply appropriate eligibility criteria (inclusion of control population), flawed measurement of exposure and outcome, failure to adequately control confounding and incomplete follow-up.

Two reviewers (MAS and SBM) independently assessed each paper across the bias domains with each being rated as either high risk, low risk or level of risk unclear as per GRADE recommendations.[94] Studies with high risk in one or more domains were considered high risk. Results of bias assessments are reported in **table 3.2** and **table 3.3**.

Table 3.2 Risk of bias (randomised controlled trials)

					RCT bias assessment						
Author, year	Design	Total Patients	Population	Industry Funding	Allocation: Generation	Allocation: Concealment	Blinding: Participants	Blinding: Assessors	Outcome: Complete	Outcome: Selective	Other Bias
Chamberlain [55]	RCT	198	Adult OOH sepsis	no	Unclear	Unclear	Unclear	Unclear	Low	Unclear	Unclear

RCT – randomised controlled trial, OOH – out of hospital

High	Significant risk of bias
Unclear	Risk of bias unclear
Low	Minimal risk of bias

Table 3.3 Risk of bias (non-randomised studies)

Author, year	Design	Total Patients	Population	Industry Funding	Non-RCT bias assessment			
					Eligibility Criteria	Exposure/ Outcome	Confounding	Follow up
Baez, *et al.* [101]	Non-RCT	63	ED sepsis patients arriving by EMS	no	Low	Low	High	Unclear
Band, *et al.* [53]	Non-RCT	963	OOH sepsis EMS vs POV	no	Low	Low	Low	Low
Bayer, *et al.* [102]	Non-RCT	56	OOH sepsis patients treated by Dr	no	Unclear	High	High	Low
Bayer, *et al.* [102]	Non-RCT	375	ED medical patients arriving by EMS	no	Low	Low	Low	Low
Erwin, *et al.* [103]	Non-RCT	151	Patients from 3 nursing homes	no	High	High	High	Unclear
Femling, *et al.* [104]	Non-RCT	151	OOH sepsis EMS vs POV	no	Low	Unclear	High	Low
Guerra, *et al.* [23]	Non-RCT	112	ED sepsis patients arriving by EMS	no	Low	Low	High	High
Hokanson, *et al.* [46]	Non-RCT	42	ED sepsis patients arriving by EMS	no	Low	High	High	Unclear
McClelland, *et al* [33]	Non-RCT	49	ED sepsis patients arriving by EMS	no	Unclear	Unclear	High	Unclear

Milzman, *et al.* [105]	Non-RCT	5,182	OOH suspected infection EMS vs POV	no	Unclear	Unclear	High	Low
Polito, *et al.* [106]	Non-RCT	66439	ED medical patients arriving by EMS	no	Low	Low	Low	Low
Seymour, *et al.* [107]	Non-RCT	144,913	ED medical patients arriving by EMS	no	Low	Low	Low	Low
Seymour, *et al.* [107]	Non-RCT	216	ED medical patients arriving by EMS	no	Low	Low	Unclear	Low
Seymour, *et al.* [25]	Non-RCT	407,176	ED medical patients arriving by EMS	no	Low	Low	Low	Low
Seymour, *et al.* [56]	Non-RCT	1350	ED severe sepsis arriving by EMS	no	Low	Low	Unclear	Low
Shiuh, *et al.* [34]	Non-RCT	183	OOH sepsis identified by sepsis protocol	no	Low	High	Unclear	Low
Studnek, *et al.* [27]	Non-RCT	311	ED severe sepsis arriving by EMS	no	Low	Unclear	Low	Low
Suffoletto, *et al.* [24]	Non-RCT	199	EMS infection patients	no	High	High	High	Low
Travers, *et al.* [35]	Non-RCT	629	OOH Sepsis patients	no	Low	Unclear	Low	Unclear
Wallgren, *et al.* [36]	Non-RCT	353	OOH sepsis	no	Low	High	High	Low

Non-RCT – non randomised controlled trial, ED – Emergency Department, EMS – Emergency Medical Services, OOH – out of hospital, POV – privately owned vehicle

High	Significant risk of bias
Unclear	Risk of bias unclear
Low	Minimal risk of bias

3.4.4 Incidence of sepsis in the prehospital environment

Seven studies addressing incidence of sepsis in the prehospital environment were identified.[24 25 32 33 56 107 108] All seven studies were non-randomised trials. The level of evidence across the studies was downgraded from LOW quality to VERY LOW quality due to risk of bias, inconsistency, indirectness, imprecision and publication bias (see **table 3.4**). Two of the studies, Seymour *et al*[107] and Seymour *et al*[25] reported data from a single large registry spanning the period 2000-2009.

Seymour *et al*[107] identified patients with critical illness, defined as patients with sepsis or patients in need of mechanical ventilation or patients who would die in hospital. They reported the incidence of critical illness to be 5.49%. However, for the subset of patients with severe sepsis incidence was 3.4%. In their later paper, Seymour *et al* reported the incidence of severe sepsis or septic shock to be 3.3 per 100 ambulance transports[25] The authors however also identify that the incidence of sepsis in the prehospital environment was steadily rising at 11.8% per year over the period of data collection (95% CI 10% - 13.7%; p<0.01). At the start of data collection in 2000 the incidence was 2.2 per hundred transports whereas in 2009 the incidence had risen to 4.2 per hundred transports (95%CI 4.1 - 4.3).[25] It was not established if the increase in incidence was secondary to increasing awareness among EMS providers, improved diagnostic capability within EMS, improved documentation at hospital or a genuine increase in disease prevalence. In their 2014 paper addressing the impact of prehospital fluid therapy on outcomes from severe sepsis they report that 1,350 of 45,394 EMS encounters (2.97%) met criteria for severe sepsis on admission to hospital.[56]

In 2013 Guerra *et al*[32] published their analysis of 15,338 EMS patients reporting that 1069 patients had infection (6.9%), 112 (10.5%) of whom were identified as having sepsis by using a sepsis screening tool. These data establish the incidence of sepsis at 112/15338 (0.7%) in this study. A similar low incidence was reported by Polito *et al*[108] who identified that 555/66,439 (0.8%) of EMS patients met their 'at risk of sepsis' criteria (heart rate >90bpm, respiratory rate >20bpm, systolic blood pressure <110mmHg) among patients with infection. In this study, the incidence of severe sepsis was reported to be 75/66,439 (0.001%). Both of these studies[32 108] suggest the incidence of sepsis may, in fact, be far lower than that previously reported by Seymour *et al*[25 56 107] A small observational study by Suffoletto *et al*[24] reported that 31/199 (15.6%) of EMS patients had severe infection, while 16/199 (8%) had severe sepsis. However, the study population

excluded trauma and stroke patients and as such these data will overestimate the true incidence of sepsis.

McClelland *et al*[33] were the only authors to publish data concerning sepsis in a UK ambulance service. They reported the number of sepsis cases within a defined period, but failed to report the total number of cases from which the sample was drawn. Estimation of the total number of cases, based upon annual activity figures, suggests a crude incidence of sepsis in a UK ambulance service in the region of 1.8%.

There is considerable variation in the reported incidence of sepsis within EMS. Documented incidence of severe sepsis range from 0.001% to 3.4%. Most data identified relate to the United States. No reliable data addressing incidence of sepsis in the United Kingdom, or elsewhere, were identified.

Table 3.4 Summary of findings (Incidence of sepsis)

№ of studies	№ of patients	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Level of evidence	Findings
Incidence of sepsis									
6	1,927	non-RCT	very serious[1]	none	not serious[2]	very serious[3]	very serious[4]	⊕◯◯◯ **VERY LOW**	**Seymour *et al* (2010a)** Incidence of severe sepsis 3.4% (4,896/144,913) **Suffoletto *et al* (2011)** 16.1% (31/199) of patients (adult, non-trauma, non-stroke) had serious infection, 50% (16/31) of infected had sepsis. Incidence among non-trauma, non-stroke patients 8% (16/199) **Seymour *et al* (2012)** 3.3% (13,249/407,176) of patients (adult, non-trauma, non-cardiac arrest) had ICD codes for severe sepsis. **Seymour *et al* (2014)** 2.97% (1350/45,394) of EMS patients diagnosed with severe sepsis **Guerra *et al* (2013)** 6.9% (1,069/15,338) of patients had infection, 10.5% (112/1,069) of infection patients met screening tool criteria for severe sepsis. Overall incidence 0.7% (112/15,338). **Polito *et al* (2015)** 0.8% **(**555/66,439) of all EMS encounters 'at risk', 13.5% (75/555) had severe sepsis. Overall incidence of severe sepsis 0.001% (75/66,439). **McClelland *et al* (2015)** crude estimate extrapolated from annual call volume ~1.8%

non-RCT: non-randomised controlled trial (observational study), EMS: Emergency Medical Services, ICD: International Classification of Disease.

1. See risk of bias tables
2. Single centre studies may limit generalizability.
3. Small study numbers limit precision/accuracy, failure to report confidence intervals
4. Publication bias (Guerra *et al*)

3.4.5 Predictors of sepsis in the prehospital environment

Twelve studies addressing predictors of sepsis in the prehospital environment were identified.[31-36 46 54 101 107-109] Four studies reported strength of correlation of individual predictor variables with the outcome of sepsis.[54 101 107 108] The remaining eight studies identified variables used to indicate sepsis, but did not report correlation of variables with the outcome of sepsis. All twelve studies were non-randomised controlled trials. Five of the studies were published in abstract only.[31 34 35 46 109] The level of evidence across the studies was downgraded from LOW quality to VERY LOW quality due to risk of bias, inconsistency, indirectness and imprecision (see **table 3.5**). The majority of the studies base their diagnostic criteria on Systemic Inflammatory Response Syndrome (SIRS) criteria as defined by the Survive Sepsis Campaign.[15]

Table 3.5 Summary of findings (predictors of sepsis)

№ of studies	№ of patients	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Level of evidence	Findings
Studies reporting correlation of variables with sepsis									
4	211790	non-RCT	very serious[1]	none	not serious[2]	very serious[3]	none	⊕◯◯◯ **VERY LOW**	**Seymour *et al* (2010a)** *Age* (Age 45-64: ß$_i$=0.91 (95% CI 0.80-1.02), Age>64: ß$_i$=1.32 (95% CI 1.22-1.43)), *Respiratory rate* (RR <12: ß$_i$=1.35 (95% CI 1.22-1.43), RR 24-35: ß$_i$=0.79 (95% CI 0.72-0.86), RR>35: ß$_i$=1.54 (95% CI 1.43-1.64)), *Systolic blood pressure* (SBP<91: ß$_i$=0.92 (95%CI 0.82 - 1.0)) , *Heart rate* (pulse>119: ß$_i$=0.77 (95%CI 0.68-0.85)), *Pulse oximetry* (SpO_2 80-87: ß$_i$=0.83 (95%CI 0.61-1.04), SpO_2<80: ß$_i$=0.1.08 (95%CI 0.82-1.35)), *Glasgow Coma Score* (GCS 12-14: ß$_i$=0.51 (95% CI 0.38-0.63), GCS 8-11: ß$_i$=1.24 (95% CI 1.10-1.39), GCS<8: ß$_i$=1.96 (95% CI 1.81-2.10)) **Baez *et al* (2013)** *RR*>20 (OR 4.81 (95%CI 1.16-21.01)) and *shock index*>0.7 (OR 5.96 (95%CI 1.49-25.78)) **Bayer *et al* (2015)** *Temperature* (Temp<36C: ß$_i$=0.98 (95%CI —0.19-2.1), Temp>38C: ß$_i$=3.92 (95%CI 3.01-4.92)), *Pulse oximetry* (SpO_2<92: ß$_i$=1.85 (95%CI 1.1-2.64)), *Respiratory rate* (RR >22: ß$_i$=1.41 (95% CI 0.62-2.26), *Heart rate* (pulse>90: ß$_i$=0.1.9 (95%CI 0.56-3.23)). **Polito *et al* (2015)** *Age* (age 50-59: Odds Ratio (OR) 3.83 (95%CI 1.05-14.07)), age>59: OR 1.63 (95%CI 0.39-6.75)), *Nursing home resident* (OR 4.47 (95%CI 1.77-11.25)), *EMS dispatch complaint of "sick person"* (OR 2.46 (95% CI 1.12-5.40)), *Hot tactile temperature (Y/N)* (Odds Ratio (OR) 2.52 (95%CI 1.10-5.74)), *Systolic blood pressure* (SBP per 1mmHg increase OR: 0.96 (95% CI 0.94-0.99)), *Oxygen saturation* (SpO_2 per 1% increase: OR 0.94 (95% CI 0.90-0.99)).

Studies identifying variables used to diagnose sepsis without reporting correlations								
8	55398	non-RCT	very serious[1]	none	not serious[2]	very serious[3]	very serious[4]	⊕◯◯◯ **VERY LOW**	**Erwin *et al* (2011)** Respiratory rate, Heart rate, Temperature, Level of consciousness, Lactate, Skin, Capillary bed refill time. **Shiuh *et al* (2012)** Respiratory rate, Heart rate, Temperature, Lactate. **Hokansen *et al* (2012)** Pulse rate, Respiratory rate, Temperature, Lactate. **Milzman *et al* (2012)** Temperature, Shock index. **Guerra *et al* (2013)** Respiratory rate, Heart rate, Temperature, Systolic blood pressure, Lactate. **Wallgren *et al* (2014)** (Robson tool) Respiratory rate, Heart rate, Temperature, Level of consciousness, Blood glucose, Skin. (BAS 90-30-90) Systolic blood pressure, respiratory rate, Oxygen saturations. **McClelland *et al* (2015)** (modified Robson tool) Respiratory rate, Oxygen saturations, Heart rate, Temperature, Level of consciousness, Blood glucose, Skin. **Bayer *et al* (2015)** (MEWS score) Respiratory rate, Heart rate, Temperature, Level of consciousness, Systolic blood pressure, Urine output.

non-RCT: non-randomised controlled trial, $ß_i$: regression coefficient, CI: confidence interval, RR: respiratory rate, SBP: systolic blood pressure, SpO2: peripheral oxygen saturations, GCS: Glasgow Coma Score, OR: Odds ratio, EMS: Emergency Medical Services, MEWS: Modified Early Warning Score

.

1. See risk of bias tables
2. Single centre studies may limit generalizability.
3. Small study numbers limit precision/accuracy
4. Abstract publication, publication bias (Guerra *et al*)

Studies reporting correlation of variables with sepsis

VERY LOW quality evidence, from four non-randomised trials,[54 101 107 108] reporting correlation of variables with sepsis was identified (see **table 3.5**). Seymour *et al*[107] developed the Critical Illness Score (CIS) to identify EMS patients who would benefit from transport to a specialist centre. They defined critical illness as severe sepsis, delivery of mechanical ventilation or death at any point during hospitalisation. Their tool was not designed to identify sepsis patients specifically. However, the authors reported operating characteristics of the tool for sepsis patients as a distinct population. The study utilised clinical records of 144913 EMS patients. Variables correlated with critical illness included age (Age 45-64: $ß_i$=0.91 (95% CI 0.80-1.02), Age>64: $ß_i$=1.32 (95% CI 1.22-1.43)), respiratory rate(RR) (RR<12: $ß_i$=1.35 (95% CI 1.22-1.43), RR 24-35: $ß_i$=0.79 (95% CI 0.72-0.86), RR>35: $ß_i$=1.54 (95% CI 1.43-1.64)), systolic blood pressure (SBP<91: $ß_i$=0.92 (95%CI 0.82 - 1.0)) , heart rate (pulse>119: $ß_i$=0.77 (95%CI 0.68-0.85)), pulse oximetry (SpO_2 80-87: $ß_i$=0.83 (95%CI 0.61-1.04), SpO_2<80: $ß_i$=0.1.08 (95%CI 0.82-1.35)), Glasgow Coma Score (GCS 12-14: $ß_i$=0.51 (95% CI 0.38-0.63), GCS 8-11: $ß_i$=1.24 (95% CI 1.10-1.39), GCS<8: $ß_i$=1.96 (95% CI 1.81-2.10)).

Bayer *et al*[54] derived the Prehospital Early Sepsis Detection (PreSep) score from a sample of 375 EMS patients, of whom 93 had sepsis (including 60 patients with severe sepsis and 12 patients with septic shock). Variables correlated with sepsis included temperature (Temp<36°C: $ß_i$=0.98 (95%CI -0.19-2.1), Temp>38°C: $ß_i$=3.92 (95%CI 3.01-4.92)), pulse oximetry (SpO_2<92: $ß_i$=1.85 (95%CI 1.1-2.64)), respiratory rate (RR >22: $ß_i$=1.41 (95% CI 0.62-2.26) and heart rate (pulse>90: $ß_i$=0.1.9 (95%CI 0.56-3.23)).

Polito *et al*[108] derived the Prehospital Recognition of Severe Sepsis (PreSS) score from a population 66,439 EMS encounters. The sample studied included 555 patients at risk of sepsis, of whom 75 were noted to have severe sepsis. Variables correlated with sepsis included age (age 50-59: Odds Ratio (OR) 3.83 (95%CI 1.05-14.07)), age>59: OR 1.63 (95%CI 0.39-6.75)), nursing home resident (OR 4.47 (95%CI 1.77-11.25)), EMS dispatch complaint of “sick person” (OR 2.46 (95% CI 1.12-5.40)), hot tactile temperature (Y/N) (Odds Ratio (OR) 2.52 (95%CI 1.10-5.74)), systolic blood pressure (SBP per 1mmHg increase OR: 0.96 (95% CI 0.94-0.99)) and oxygen saturation (SpO_2 per 1% increase: OR 0.94 (95% CI 0.90-0.99)).

Baez *et al*[101] completed an observational study of adult EMS patients with a diagnosis of sepsis transported to hospital. Their population comprised 63 patients only, so statistical estimates should be interpreted with caution because of high uncertainty. Among patients with sepsis the only prehospital physiological variables associated with admission to the Intensive Care Unit were respiratory rate>20 (OR 4.81 (95%CI 1.16-21.01)) and shock index>0.7 (OR 5.96 (95%CI 1.49-25.78)).

The preceding four studies were the only studies to report statistical estimates addressing correlation of variables with sepsis. The remaining studies identified variables and any intervals or thresholds, but failed to report any statistical estimates for the individual variables.

Studies identifying variables used to diagnose sepsis without reporting correlations

VERY LOW quality evidence, from eight non-randomised trials,[31-34 36 46 54 109] identifying variables used to diagnose sepsis was identified (see **table 3.5**). In their prospective observational study Erwin *et al*[31] defined sepsis as "identified or suspected infection, plus two of abnormal temperature (Temp<96.8° or >100.4° F), heart rate > 90/min, respiratory rate>20/min or altered mental status" (altered mental status was not explicitly defined). Severe sepsis was defined as "Sepsis plus mottled skin or capillary refill>3 seconds or lactate> 2mmol/L or abrupt changes in mental status". Guerra *et al*[32] defined their sepsis criteria as 18 years or older and not pregnant, plus suspected or documented infection, plus two systemic inflammatory response syndrome criteria (temperature >38° or <36°C, pulse > 90 bpm, respiratory rate > 20 bpm or mechanically ventilated) plus hypoperfusion (systolic blood pressure < 90mmHg or mean arterial pressure or 65mmHg, lactate > 4 mmol/L).

Wallgren *et al*[36] compared Robson Tool and the BAS 90-30-90 score with EMS clinician judgement to identify sepsis patients. The Robson Tool defines sepsis as signs and symptoms of infection, plus any two of temperature > 38.3° or < 36°C, heart rate > 90 bpm, respiratory rate > 20 bpm and acutely altered mental status, plasma glucose > 6.6mmol/L (unless diabetic). The BAS 90-30-90 employs thresholds for sepsis at systolic blood pressure below 90 mmHg, respiratory rate > 30 bpm and oxygen saturations below 90%. McClelland *et al*[33] report use of a modified Robson tool, whereby an additional variable, pulse oximetry (SpO_2<90%), was included over and above those previously identified.

Three studies were reported in abstract form only.[34 46 109] Shiuh *et al*[34] defined sepsis as the presence of "2 or more systemic inflammatory response syndrome (SIRS) criteria (continuous pulse rate>90, Respiratory Rate >20 and Temperature >38 or <36 C) plus clinical suspicion for infection". In this study, EMS crews also stratified sepsis patients according to prehospital lactate readings, if patients had a lactate >4mmol/L paramedic crews provided the hospital with an '*alert*' message whereas if the lactate was in the range 2.5-3.9 mmol/L they provided the hospital with an '*inform*' message prior to, or on hospital arrival. Hokansen *et al*[46] defined sepsis as 2 or more SIRS criteria without defining the SIRS criteria, and in addition utilised lactate to stratify risk. Milzman *et al*[109] described the utility of temperature and shock index as a pair of observations to predict sepsis.

One study[54] included an assessment of the Modified Early Warning Score (MEWS) to identify patients with sepsis. MEWS is a generic score that was proposed by Stenhouse *et al*[110] and is used to estimate the likelihood of patient deterioration, needing subsequent surgical intervention or admission to intensive care. It is a physiological scoring system that can be used by any healthcare provider (see **figure 3.2**). It is intended to be used at regular intervals, rather than at initial assessment only, to identify a deteriorating patient and trigger early intervention by the medical team.

Figure 3.2 Modified Early Warning Score (MEWS)

MEWS (Modified Early Warning System)

	3	2	1	0	1	2	3
Respiratory Rate per minute		Less than 8		9-14	15-20	21-29	More than 30
Heart Rate per minute		Less than 40	40-50	51-100	101-110	111-129	More than 129
Systolic Blood Pressure	Less than 70	71-80	81-100	101-199		More than 200	
Conscious level (AVPU)	**U**nresponsive	Responds to **P**ain	Responds to **V**oice	**A**lert	New agitation Confusion		
Temperature (°c)		Less than 35.0	35.1-36	36.1-38	38.1-38.5	More than 38.6	
Hourly Urine For 2 hours	Less than 10mls / hr	Less than 30mls / hr	Less than 45mls / hr				

EARLY WARNING SCORING SYSTEM FOR DETECTING ADULT PATIENTS WHO HAVE OR ARE DEVELOPING CRITICAL ILLNESS
IS THE SCORE FOR YOUR PATIENT 1-2? PERFORM 2 HOURLY OBSERVATIONS AND INFORM NURSE IN CHARGE
IS THE SCORE FOR YOUR PATIENT 3? PERFORM 1-2 HOURLY OBSERVATIONS AND INFORM NURSE IN CHARGE
*IF THE MEWS SCORE IS DETERIORATING : THE WARD S.H.O. OR DUTY DOCTOR **MUST** ATTEND*
IS THE SCORE FOR YOUR PATIENT 4 OR MORE? PERFORM OBSERVATIONS AT LEAST 1/2 HOURLY. ENSURE MEDICAL ADVICE IS SOUGHT AND CONTACT OUTREACH TEAM (see below)

(Image from https://twitter.com/blairbigham/status/823956305266311168)

Variables that have been used to inform sepsis recognition across the studies identified are summarised in **table 3.6**.

Table 3.6 Variables used to identify sepsis

	Variable														
Screening tool (or study author)	Respiratory rate[α]	Heart rate[α]	Temperature[α]	LOC[β]	SpO_2[β]	Blood pressure[β]	Shock index	Lactate[β]	Blood -	Skin	CBRT	Dispatch category	Location	Age	Urine output
Seymour *et al* (CIS)	•	•		•	•	•								•	
Polito *et al* (PRESS)			•		•	•						•	•	•	
Bayer *et al* (PRESEP)	•	•	•		•	•									
Baez *et al*	•						•								
Wallgren *et al* (Robson)	•	•	•	•				•	•	•					
McClelland (modified Robson)	•	•	•	•	•			•	•	•					
Bayer *et al* (BAS 90-30-90)	•				•	•									
Bayer *et al* (Modified Early Warning Score (MEWS))	•	•	•	•		•									•
Erwin *et al*	•	•	•	•				•		•	•				
Guerra *et al*	•	•	•			•		•							
Hokansen *et al*	•	•	•					•							
Milzman *et al*			•				•								
Shiuh *et al*	•	•	•					•							

[α] Systemic Inflammatory Response Syndrome (SIRS) criteria, [β] organ dysfunction, LOC reduced level of consciousness, SpO_2 oxygen saturations, CBRT capillary bed refill time

3.4.6 Accuracy of prehospital sepsis screening

Nine studies describing diagnosis of sepsis among patients in the prehospital environment were identified.[31-36 54 107 108] All nine studies were non-randomised controlled trials. Three were published in abstract only.[31 34 35] The level of evidence across the studies was downgraded from LOW quality to VERY LOW quality due to risk of bias, inconsistency, indirectness, imprecision and publication bias (see **table 3.7**).

Diagnostic accuracy is not simply concerned with the proportion of patients identified as having sepsis (true positives) versus the proportion of patients incorrectly identified as septic (false positives). Measures of diagnostic accuracy

should also consider those patients correctly identified to not be septic (true negatives) as well as those incorrectly identified to not be septic (false negatives).

For the bedside clinician to assess diagnostic accuracy of a screening tool it is vital that they can interpret a test result and apply it to their clinical practice. In other words, if the screening tool suggests the patient has the disease, how likely is it that the patient actually has disease? For this, positive predictive values (PPV) are important. Conversely, if screening suggests the patient does not have the disease how likely is it that the patient is actually disease free? Negative predictive values (NPV) will inform the clinician in this instance. Screening tool performance can be assessed by reporting the following measures:

- Sensitivity - the ability of a test to identify those with the disease. A test with sensitivity = 1 identifies all cases with the disease.
- Specificity - the ability of a test to identify those who do not have the disease. A test with specificity = 1 identifies all cases without the disease.
- Positive predictive value (PPV) - probability that an individual with a positive test has the disease. A test with PPV = 1 implies all cases classified as having disease will have the disease.
- Negative predictive value (NPV) - probability that an individual with a negative test does not have the disease. A test with NPV = 1 implies all cases classified as being disease free will not have the disease.

Table 3.7 - Summary of findings (Accuracy of sepsis diagnosis)

№ of studies	№ of patients	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Level of evidence	Findings
Accuracy of prehospital sepsis screening tools (not clinically validated by EMS)									
3	211727	non-RCT	none	none	not serious[1]	very serious[2]	none	⊕◯◯◯ **VERY LOW**	**Seymour (2010)** Critical Illness Score (CIS): Risk of Sepsis 0.76 (95%CI 0.75-0.77). **Polito (2015)** PreSS score: Sensitivity 0.85, specificity 0.47, PPV 0.19, NPV 0.96 (95% CI not reported) PreSS score >=3 sensitivity 0.81 specificity 0.63 **Bayer (2015)** PreSep score: Sensitivity 0.85 (95% CI 0.77 to 0.92), specificity 0.86 (95% CI 0.82 to 0.90), PPV 0.66, NPV 0.95
Accuracy of prehospital sepsis screening tools (in clinical use by EMS)									
2	161	non-RCT	very serious[3]	none	not serious[4]	very serious[5]	very serious[6]	⊕◯◯◯ **VERY LOW**	**Guerra (2013)** Screening based on SSC criteria identified 32/67 sepsis patients (47.8%). (95% CI not reported). **McClelland (2015)** Screening using modified Robson tool. Sensitivity & Specificity for sepsis 43% (95%CI 28-58%) & 14% (95%CI 0-40%) respectively. Sensitivity and specificity for severe sepsis 30% (95%CI 12%-47%) & 77% (95%CI 60-95%).

Retrospective application of EMS data to screening tool by researcher								
2	728	non-RCT	very serious[3]	none	not serious[4]	very serious[5]	none	⊕◯◯◯ **VERY LOW**	**Wallgren (2014)** Retrospective application of two different screening tools in comparison to clinical judgement. For sepsis Robson tool: sensitivity 75% (p<0.001), BAS 90-30-90: sensitivity 43% (p<0.001), Clinical judgement: 12% accuracy. (95% CI not reported). For severe sepsis Robson tool: sensitivity 93% (p<0.001), BAS 90-30-90: sensitivity 70% (p<0.001), Clinical judgement: 17% accuracy. (95% CI not reported). **Bayer (2015)** Retrospective application of three different screening tools. (modified) Robson tool: sensitivity 0.95, specificity 0.43, PPV 0.32, NPV 0.97. BAS 90-30-90: sensitivity 0.62, specificity 0.83, PPV 0.51, NPV 0.89. MEWS≥4 sensitivity 0.74, specificity 0.75, PPV 0.45, NPV 0.91. (95% CI not reported).
Sepsis recognition by EMS without using a screening tool.								
4	1316	non-RCT	very serious[3]	none	not serious[4]	very serious[5]	very serious[7]	⊕◯◯◯ **VERY LOW**	**Erwin (2011)** Screening based on SSC criteria. For sepsis: sensitivity 33% (95%CI 18-53%), specificity 89% (95%CI 08-94%), PPV 50% (95%CI 28-72%), NPV 80% (95%CI 70-87%). For severe sepsis: sensitivity 20% (95%CI 5-51%), specificity 94% (95%CI 87-97%), PPV 29% (95%CI 08-64%), NPV 91 (95%CI 83-95%). **Shiuh (2012)** Screening based on SSC criteria, also stratified by lactate, lactate<4 = "sepsis advisory" while lactate>4 = "sepsis alert". 74.2% of "Sepsis Advisory" patients and 76.7% of "Sepsis Alert" patients received a hospital diagnosis of severe infection or sepsis. (95% CI not reported). **Travers (2013)** Screening criteria not defined. Specificity 78.85% (95% CI 75.23-82.17), sensitivity 73.4% (95% CI 61.40-83.05), PPV 30.59% (95% CI 23.76-38.11), NPV 95.86% (95% CI 93.61-97.49), accuracy 78% (52 true positives, 440 true negatives). **Wallgren (2014)** For sepsis clinical judgement: 12% accuracy (95% CI not reported). For severe sepsis clinical judgement: 17% accuracy (95% CI not reported).

non-RCT: non-randomised study; SSC: Surviving Sepsis Campaign; PPV: positive predictive value; NPV: negative predictive value; BAS 90-30-90 = SBP < 90 mm Hg; respiratory rate> 30 breaths/min; SpO_2 < 90%; MEWS: modified early warning score; Modified Robson = Robson tool with addition of SpO_2

1. Seymour *et al* CIS not specific to sepsis (CIS intended to identify all cases of critical illness). Polito *et al* and Bayer *et al* studies limited to single EMS systems, Bayer *et al* physician based EMS
2. Polito *et al* failed to report confidence intervals, small sample size in Bayer et al study
3. See bias assessments
4. Single EMS agency/hospital so limited generalisablity.
5. Included studies have small sample sizes thus imprecise point estimates. In several studies confidence intervals are not reported.
6. Guerra *et al* publication bias likely.
7. Studies published in abstract only, unable to reliably critically appraise.

Accuracy of prehospital sepsis screening tools (not clinically validated by EMS)

VERY LOW quality evidence was identified, from three non-randomised trials,[54] [107] [108] addressing accuracy of prehospital sepsis screening tools (see **table 3.7**). Each of the studies adopted a similar approach to screening tool development. Identification candidate predictors varied slightly between studies. However, once candidate predictors were identified all studies utilised multivariable logistic regression, in a stepwise fashion, to build their respective models. Goodness of fit was assessed by Hosmer-Lemeshow test and model performance determined by calculating the area under the receiver operating characteristic curve. None of the studies included a clinical validation study of their respective screening tools.

Seymour *et al*[107] developed the Critical Illness Score (CIS) to predict the risk of critical illness among EMS patients. It was not developed to identify sepsis specifically, although the statistical estimates reported in this review relate to sepsis cases only. Their study utilised the clinical records of 144,913 EMS patients 4,895 of whom had severe sepsis. Polito *et al*[108] derived the PreSS score from a population 66,439 EMS encounters. The study included 555 patients at risk of sepsis, of whom 75 had severe sepsis. Bayer *et al*[54] derived the Prehospital Early Sepsis Detection (PreSep) score from a sample of 375 EMS patients, of whom 93 had sepsis (including 60 patients with severe sepsis and 12 patients with septic shock).

CIS, PRESS and PreSEp were derived and validated with EMS data. Reported operating characteristics do not relate to use in clinical practice by ambulance clinicians.

Accuracy of screening tools (in clinical use by EMS)

VERY LOW quality evidence was identifed, from two non-randomised trials,[32] [33] addressing recognition of sepsis by EMS personnel utilising a screening tool (see **table 3.7**). Guerra *et al*[32] report that Emergency Medical Technicians (EMTs) trained to recognise sepsis correctly identified 32/67 (sensitivity 48%) septic patients, with failure to recognise sepsis in 35/67 (False Negative Rate (FNR) 52%) of cases. However, the FNR may be misleading. In 5/35 (14%) of cases the patient's vital signs did not meet the criteria of the sepsis screening tool while in EMS care, in 8/35 (23%) of cases the patients had cryptic shock but EMTs did not have lactate meters to enable detection and in 13/35 (37%) of cases

diagnosis was made by abnormal white cell count (only available in hospital). In 9/35 (26%) of cases EMTs failed to identify sepsis when sufficient diagnostic criteria were available to them. The high proportion of patients missed due to lack of white cell count highlights a limitation of prehospital sepsis screening tools. Guerra *et al*[32] further reported that among sepsis patients transported by EMS crews not trained to recognise sepsis 5/45 (11%) were identified as septic.

McClelland and Jones[33] scrutinised the records of all septic patients conveyed by a regional ambulance service to a university hospital to determine if ambulance clinicians (previously trained in the use of a screening tool) recognised and documented suspected sepsis. The screening tool used was based upon the Robson tool amended to include oxygen saturations as an indicator of organ dysfunction.[33] McClelland and Jones[33] report a sensitivity of 0.43 (95% CI 0.28-0.58) for sepsis and 0.30 (95% CI 0.12-0.47) for severe sepsis and a specificity of 0.14 (95% CI 0.0-0.40) for sepsis and 0.77 (95% CI 0.60-0.95) for severe sepsis. The authors concluded that use of the screening tool by ambulance clinicians was inconsistent but that it improved sepsis recognition.[33]

Retrospective application of EMS data to screening tool by researcher

VERY LOW quality evidence was identified, from two non-randomised trials,[36 54] addressing retrospective application of prehospital data to screening tools (see **table 3.7**). Wallgren *et al*[36] compared two screening tools (Robson Tool and BAS 90-30-90 score) with EMS clinician judgement. The Robson Tool performed better than the BAS 90-30-90 score (see **table 3.7**). Sensitivity with the Robson tool was 0.75 (95% CI not reported) and 0.93 (95% CI not reported) for sepsis and severe sepsis respectively. Sensitivity with the BAS 90-30-90 tool was 0.43 (95% CI not reported) and 0.93 (95% CI not reported) for sepsis and severe sepsis respectively.

Bayer *et al*[54] compared the performance of their PRESEP score with the Modified Early Warning Score (MEWS), BAS 90-30-90 and Robson tool reporting that the PRESEP score surpassed both MEWS and BAS 90-30-90 for sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The Robson tool showed better sensitivity however the PRESEP tool had better specificity. Furthermore, the PRESEP score showed better PPV and comparable NPV than the Robson tool (see **table 3.7**).

Sepsis recognition by EMS without using a screening tool.

VERY LOW quality evidence was identified, from three non-randomised trials,[31] [34] [35] addressing accuracy of paramedic diagnosis of sepsis in clinical practice. All three studies were published in abstract form only (see **table 3.7**). Erwin *et al*[31] compared paramedic diagnosis of sepsis and severe sepsis with physician diagnosis. The level of agreement between paramedics and physicians was low (kappa=0.25 and 0.16 respectively). These results lead the authors to conclude that sepsis criteria were more useful for ruling-out sepsis than diagnosing sepsis. In a study by Shiuh *et al*[34] EMS crews stratified sepsis patients according to prehospital lactate readings. If patients had a lactate>4mmol/L paramedic crews provided the hospital with an '*alert*' message whereas if the lactate was in the range 2.5-3.9 mmol/L they provided the hospital with an '*inform'* message prior to, or on hospital arrival. They reported data for 219 sepsis patients for whom a lactate reading was available. They did not report data for those patients where a lactate reading was not known/unavailable. Travers *et al*[35] compared sepsis diagnosis by paramedics and ED doctors in 629 cases. Diagnosis by the ED doctor was deemed the diagnostic gold standard. However, it was a subjective diagnosis and was not verified against any established diagnostic criteria. Thermometry was not available to paramedics to accurately determine body temperature. Paramedic diagnosis was reported to be consistent with physician diagnosis in 78% of cases. This is the largest paramedic diagnostic accuracy study but unfortunately detail is lacking. Finally, Wallgren *et al*[36] reported that paramedics documented clinical suspicion of sepsis, defined as "documentation of suspected sepsis, septicaemia, urosepsis or blood poisoning in the patient's clinical record", in 11.9% and 16.9% of cases for sepsis and severe sepsis respectively. Confidence intervals were not reported.

Accuracy of prehospital sepsis diagnosis across the studies identified is summarised in **table 3.8**.

Table 3.8 Performance of screening tools

Author	Sensitivity	Specificity	PPV	NPV
Seymour (CIS)	0.76 (95%CI 0.75-0.77)	Not reported	Not reported	Not reported
Polito (PreSS)	0.85 (95% CI not reported)	0.47 (95% CI not reported)	0.19 (95% CI not reported)	0.96 (95% CI not reported)
Bayer (PreSep)	0.85 (95% CI 0.77-0.92)	0.86 (95% CI 0.82-0.90)	0.63 (95% CI not reported)	0.95 (95% CI not reported)
McClelland (sepsis) (modified Robson Tool)	0.43 (95%CI 0.28-0.58)	0.14 (95%CI 0-0.40)	Not reported	Not reported
McClelland (severe sepsis) (modified Robson Tool	0.30 (95%CI 0.12-0.47)	0.77 (95% CI 0.60-0.95)	Not reported	Not reported
Bayer (modified Robson Tool)	0.95 (95% CI not reported)	0.43 (95% CI not reported)	0.32 (95% CI not reported)	0.97 (95% CI not reported)
Wallgren (sepsis) (Robson Tool)	0.75 (95% CI not reported)	Not reported	Not reported	Not reported
Wallgren (severe sepsis) (Robson Tool)	0.93 (95% CI not reported)	Not reported	Not reported	Not reported
Bayer (BAS 90-30-90)	0.62 (95% CI not reported)	0.83 (95% CI not reported)	0.51 (95% CI not reported)	0.89 (95% CI not reported)
Wallgren (sepsis) (BAS 90-30-90)	0.43 (95% CI not reported)	Not reported	Not reported	Not reported
Wallgren (severe sepsis) (BAS 90-30-90)	0.70 (95% CI not reported)	Not reported	Not reported	Not reported
Bayer (MEWS)	0.74 (95% CI not reported)	0.75 (95% CI not reported)	0.45 (95% CI not reported)	0.91 (95% CI not reported)
Guerra	0.48 (95% CI not reported)	Not reported	Not reported	Not reported
Erwin (sepsis)	0.33 (95%CI 0.18-0.53)	0.89 (95%CI 0.08-0.94)	0.50 (95%CI 0.28-0.72)	0.80 (95%CI 0.70-0.87)
Erwin (severe sepsis)	0.20 (95%CI 0.05-0.51)	0.94 (95%CI 0.87-0.97)	0.29 (95%CI 0.08-0.64)	0.91 (95%CI 0.83-0.95)
Shiuh	0.75 (95% CI not reported)	Not reported	Not reported	Not reported
Travers	0.73 (95% CI 0.61-0.83)	0.79 (95% CI 0.75-0.82)	0.31 (95% CI 0.24-0.38)	0.96 (95% CI 0.94-0.98)

CI – confidence interval

3.4.7 Impact of prehospital recognition on processes of care

Eight studies describing the impact of prehospital sepsis diagnosis, by ambulance clinicians, on processes of care were identified.[27 29 32 33 53 55 102 111] One study was a randomised controlled trial[55] while seven studies were non-randomised controlled trials.[27 29 32 33 53 102 111] Three studies[29 55 102] were published in abstract form only. The level of evidence across the studies was downgraded from LOW quality to VERY LOW quality due to risk of bias, inconsistency, indirectness, imprecision and publication bias (see **table 3.9**).

Table 3.9 Summary of findings (impact on processes of care)

№ of studies	№ of patients	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Level of evidence	Findings
Impact of prehospital care upon time to antimicrobial therapy									
1	199	RCT	not serious[1]	none	not serious[2]	very serious[3]	very serious[4]	⊕○○○ **VERY LOW**	**Chamberlain 2009** prehospital antibiotics provided 3.4 ± 2.6 hours sooner (p=0.02).
5	1,927	non-RCT	very serious[5]	none	not serious[2]	very serious[6]	very serious[7]	⊕○○○ **VERY LOW**	**Band 2011** Median time to antibiotics reduced: 116 min (IQR 66-199min) EMS vs 152 min (IQR 92-252min) 'other means' (p≤0.001). **Studnek 2012** if arriving by EMS vs other means time to antibiotics reduced 111min (EMS) vs 146min (non-EMS); (p=0.001). If EMS recognised and documented sepsis, time to antibiotics reduced 70min (documented) vs 122min (not documented) (p=0.003). **Bayer 2013** Median time of administration 19min (IQR 18-24min) after initial emergency call (time of administration estimated as 10min after arriving at scene). **Guerra 2013** No significant reduction in time to antibiotics mean 72.6min (SD 59.3min) (pre-alert) vs 98.5min (SD 89.9min) (no pre-alert) (p=0.07). **Femling 2014** Time to antibiotics: 87min (EMS) (IQR 44-157min) vs 120min (non-EMS) (IQR 141-271min), diff 33min (p=0.02).

Impact of prehospital care upon fluid resuscitation									
4	1347	non-RCT	very serious[5]	none	not serious[2]	very serious[6]	very serious[8]	⊕◯◯◯ **VERY LOW**	**Seymour 2010]** patients who received prehospital fluids had shorter time to MAP>65mmHg 17/24 (70%) (EMS IV fluids) vs 12/26 (44%) (no IV fluids), Unadj RR 1.53 (95% CI 0.9-2.65), and shorter time to CVP>8mmH_20 15/25 (60%)(EMS IV fluids) vs17/24 (70%) (no IV fluids), Unadj RR 1.2 (95% CI 0.8-1.8) **Band 2011]** Median time to initiation of IV fluid reduced: 34 min (IQR 10 - 88) EMS vs 68 min (IQR 25-121min) 'other means' of arrival (p≤0.001). **Bayer 2013]** Median time of administration 19min (IQR 18-24min) after initial emergency call (time of administration estimated as 10min after arriving at scene). Patients received 2.5l intravascular fluid (IQR 1.5–3.0l) until admitted to the ER. **Guerra 2013]** No significant difference in fluid administration by 6 hours 42.97 cc/kg (SD 33.23cc/kg) (pre-alert) vs 35.17cc/kg (SD 26.81 cc/kg) (no pre-alert) (p=0.30).

Impact of prehospital care upon Early Goal Directed Therapy									
5	1173	non-RCT	very serious[5]	none	not serious[2]	very serious[6]	very serious[8]	⊕◯◯◯ **VERY LOW**	**Seymour 2010** patients who received prehospital fluids had shorter time to MAP>65mmHg 17/24 (70%) (EMS IV Fluids) vs 12/26 (44%) (no IV fluids), Unadj RR 1.53 (95% CI 0.9-2.65); shorter time to CVP>8mmH_20 15/25 (60%) (EMS IV Fluids) vs17/24 (70%) (no IV fluids), Unadj RR 1.2 (95% CI 0.8-1.8); and shorter time to svcO2>70% 13/24 (54%) (EMS IV Fluids) vs 9/25 (36%) (no IV Fluids), Unadj RR 1.5 (95% CI 0.8-2.9). **Studnek 2012** if arriving by EMS vs other means time to EGDT reduced 119min (EMS) vs 160 min (non-EMS) (p=0.005). If EMS recognised and documented Sepsis time to EGDT 69min (documented) vs 131 min (not documented) (p=0.001). **Guerra 2013** No significant reduction in proportion of patients with central venous line placement 62% (pre-alert) vs 68% (no pre-alert) (p=0.54). **Femling 2014** Time to central line: 200min (EMS) (IQR 89-368min) vs 275min (non-EMS) (IQR 122-470min), diff 75min (p<0.01). (95% CI 0.34-0.98). **McClelland 2015** Time to 'Sepsis 6': mean 205min (SD 271min, range 10-720min)* (EMS identified) vs 120 min (SD 110, 17-450min) (not identified) * includes outlier where the fluid balance chart was not started for 12 hours, excluding this case mean 76min (SD 95min, range 10-240min)].

non-RCT: non-randomised controlled trial (observational study), IQR: inter quartile range, EMS: Emergency Medical Services, SD: standard deviation, MAP: mean arterial pressure, IV: intravascular,
RR: risk ratio, 95% CI: 95% confidence interval' CVP: central venous pressure, EGDT: early goal directed therapy.

1. Risk of bias unclear.
2. Single centre study may limit generalizability.
3. Small study numbers limits precision/accuracy.
4. Published in abstract only, insufficient detail to rule out other bias.
5. Concerns relating to eligibility, exposure, confounding, follow-up
6. Small study numbers limits precision/accuracy, failure to report confidence intervals (Guerra)
7. Abstract only publication (Femling), insufficient detail to rule out other bias, Publication bias likely (Guerra)
8. Publication bias likely (Guerra)

Time to antibiotic therapy

VERY LOW quality evidence was identified, from one randomised controlled trial[55] and five non-randomised trials,[27 29 32 53 102] addressing time to antibiotic therapy (see **table 3.9**). Three of these studies were published in abstract form only.[29 55 102]

Three studies[27 29 53] indicate that Emergency Department (ED) antibiotic therapy is administered 30-50 minutes sooner if EMS identify sepsis and inform the receiving clinician of their diagnosis. However, this finding is not universal - Guerra *et al*[32] identified a reduction in time to antibiotic therapy (pre-alert: 72.6min Standard Deviation (SD) 59.3min) vs no pre-alert: 98.5min (SD 89.9min), p=0.07), however this difference was not statistically different.

Two studies[55 102] address prehospital administration of antibiotic therapy. Chamberlain[55] reported that antibiotics were delivered 3.4±2.6 hours sooner while Bayer *et al*[102] noted that among EMS sepsis patients median time to antibiotics was 19min (IQR 18-24min) from initial emergency call. However, this should be interpreted with caution as time of administration was estimated to commence 10 minutes after arriving at scene. Furthermore, there was no indication of time to antibiotic therapy in hospital, thus it is not possible to draw conclusions as to the amount of time saved by administering antibiotics before arriving at hospital.

Time to intravascular fluid therapy

VERY LOW quality evidence was identified, from four non-randomised trials,[32 53 102 111] addressing time to intravascular therapy (see **table 3.9**). One of these studies was published in abstract only.[102]

Seymour *et al*[111] noted that patients who received prehospital fluids had shorter time to Mean Arterial Pressure (MAP)>65 mm Hg 17/24 (70%, EMS IV fluids) vs 12/26 (44%, no IV fluids), unadjusted RR 1.53 (95% CI [0.9-2.65]). Band *et al*[53] reported that arrival by EMS reduces time to initiation of intravascular fluid therapy when compared with those who arrive by privately owned vehicle (POV) (EMS: 34 min (IQR 10-88min) vs POV: 68 min, IQR 25-121min, p≤0.001). Similarly, Bayer *et al*[102] noted that among EMS sepsis patients median time to initiation of Intravascular fluids was 19min (IQR 18-24min) from initial emergency call (time of administration was estimated to commence 10 min after

arriving at scene), with patients receiving an average of 2.5l intravascular fluid (IQR 1.5–3.0l) until admission to the ED. A third study by Guerra *et al*[32] indicated that early identification of sepsis by EMS was not associated with improved 6-hour fluid resuscitation targets in the ED (EMS pre-alert: 42.97 cc/kg (SD 33.23cc/kg) vs no EMS pre-alert: 35.17cc/kg (SD 26.81 cc/kg), p=0.30).

Time to Early Goal Directed Therapy (EGDT) Targets

VERY LOW quality evidence was identified, from five non-randomised trials,[27 29 32 33 111] addressing time to Early Goal Directed Therapy targets (see **table 3.9**). One of these studies was published in abstract form only.[29]

Femling *et al*[29] reported that patients who arrived at the ED via EMS had shorter time to central line placement (required for central venous pressure monitoring) than those who arrived by other means (EMS: 200 min (IQR 89-368min) vs non-EMS:275 min (IQR 122-470 min), difference 75 min, p<0.01), while Guerra *et al*[32] noted that when EMS provided a sepsis pre-alert to the hospital the advance notification it did not impact the decision to place a central venous catheter (EMS pre-alert: 61% vs no EMS pre-alert: 68%, p=0.54). Although Seymour *et al*[111] reported that higher proportion of patients achieved a $SvcO_2$>70% within 6 hours when EMS initiated fluid therapy prior to arriving at the ED, the unadjusted Risk Ratio (RR) found no evidence of a difference (EMS IV fluids: 13/24 (54%) vs no IV fluids: 9/25 (36%), unadjusted RR 1.5, 95%CI 0.8-2.9). This same study also identified no improvement in time to MAP>65mmHg (EMS IV fluids: 17/24 (70%) vs no IV fluids: 12/26 (44%), unadjusted RR 1.53 (95% CI 0.9-2.65)), and time to CVP>8 mmH_20 (EMS IV fluids: 15/25 (60%) vs no IV fluids: 17/24 (70%), unadjusted RR 1.2 (95% CI 0.8-1.8)).[111]

Studnek *et al*[27] reported that if patients arrived by EMS they had shorter times to EGDT than if they arrived by other means (EMS: 119min vs non-EMS:160 min, SD/range not reported, p=0.005). Furthermore, among EMS transported patients, if EMS documented suspicion of sepsis then time to EGDT was shorter than if they did not document suspicion of sepsis (documented suspicion: 69min vs not documented: 131 min, SD/range not reported, p=0.001). McClelland *et al*[33] similarly reported that time to delivery of the 'Sepsis 6' (administration of supplemental oxygen, intravenous fluids, antibiotics, measurement of venous lactate, urine output, and drawing blood to identify causative pathogen) was shorter if EMS identified sepsis prior to arrival at hospital (EMS identified: mean

205min (SD 271min, range 10-720min) versus not identified: mean 120 min (SD 110, 17-450min).

3.4.8 Impact of prehospital care on clinical outcomes

Six studies describing the impact of prehospital sepsis care, by ambulance clinicians, on patient outcomes were identified.[29 32 33 53 55 56] One study was a randomised controlled trial[55] while five studies were non-randomised controlled trials.[29 32 33 53 56] Two studies were published in abstract form only.[29 55] The level of evidence across the studies was downgraded from LOW quality to VERY LOW quality due to risk of bias, inconsistency, indirectness, imprecision and publication bias (see **table 3.10**).

Impact of prehospital care on Intensive Care Unit care

VERY LOW quality evidence was identifed, from four non-randomised trials,[29 32 33 56] addressing admission to intensive care (see **table 3.10**). One of the studies was published in abstract only.[29]

Guerra *et al*[32] found that prehospital care did not lead to a statistically significant reduction in length of intensive care unit (ICU) stay: mean 7.3 days (SD 6.8 days, EMS pre-alert) vs 8.4 days (SD 8.8 days, no EMS pre-alert, p=0.65). Similarly, Femling *et al*[29] failed to identify any reduction in length of stay when patients arrived by EMS versus other means: 15 days (IQR 13-17 days, EMS) vs 14 days (IQR 10-17 days, non-EMS), confidence interval and p value not given but stated not significant. Conversely, McClelland *et al*[33] reported a large difference in admission rates when EMS staff recognised and treated patients as septic compared to when sepsis was not recognised: 4% (EMS identified) versus 13% (not identified). However, it should be noted that the sample size was very small indeed (n=23) and no confidence intervals or estimate of significance was reported. Seymour *et al*[56] reported that prehospital vascular access reduced intensive care unit admission adjusted OR 0.41 (95% CI [0.24 - 0.70]). However, prehospital intravascular fluid therapy was not associated with reduced ICU admission (adjusted OR 0.64; 95%CI 0.37-1.10).

Table 3.10 Summary of findings (impact on outcomes)

№ of studies	№ of patients	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Level of evidence	Findings
Impact of prehospital care upon ICU admission									
4	1996	non-RCT	very serious[1]	none	not serious[2]	very serious[3]	very serious[4]	⊕◯◯◯ **VERY LOW**	**[Guerra 2013]** No significant reduction in length of stay: mean 7.3 days (SD 6.8days) (Pre-alert) vs 8.4 days (SD 8.8 days) (no pre-alert) (p=0.65). **[Femling 2014]** Length of stay: 15 days (IQR13-17days) (EMS) vs 14days (IQR 10-17days) (non-EMS), diff 1 day, not significant. **[Seymour 2014]** Prehospital vascular access reduced ICU admission adjusted OR 0.41 (95% CI 0.24 - 0.70). **[McClelland 2015]** ICU admission: 4% (1/23) (EMS identified) vs 13% (3/23) (not identified).
Impact of prehospital care upon mortality									
5	2959	non-RCT	very serious[1]	none	not serious[2]	very serious[3]	very serious[4]	⊕◯◯◯ **VERY LOW**	**[Band 2011]** No significant difference in mortality was noted: adjusted RR 1.24 (95% CI 0.92 - 1.66) (p=0.16). **[Guerra 2013]** If hospital was 'pre-alerted', unadjusted mortality was improved OR 3.19 (95% CI 1.14– 8.88) (p=0.04) **[Femling 2014]** No significant difference in mortality was noted 113/378 (30%) (EMS) vs 34/107 (31%) (non-EMS), diff 1%, not significant. **[Seymour 2014]** Prehospital vascular access reduced mortality adjusted OR 0.31 (95% CI 0.17 - 0.57) (p<0.01). **[McClelland 2015]** 3 month mortality 21% (5/24) (EMS identified) vs 16% (4/25) (not identified).

Impact of prehospital antimicrobial therapy on ICU admission									
1	199	RCT	not serious[1]	none	not serious[2]	very serious[3]	very serious [5]	⊕◯◯◯ **VERY LOW**	**[Chamberlain 2009]** Mean ICU length of stay: reduced 6.8 ± 2.1 days (intervention) vs 11.2 ± 5.2 days (control) (p=0.001).
Impact of prehospital antimicrobial therapy on mortality									
1	199	RCT	not serious[1]	none	not serious[2]	very serious[3]	very serious [5]	⊕◯◯◯ **VERY LOW**	**[Chamberlain 2009]** 28-day mortality reduced: 42.4% (intervention) vs 56.7% (control), OR 0.56 (95% CI 0.32 to 1.00) (p=0.049)
Impact of prehospital intravenous fluid therapy on ICU admission									
1	1350	non-RCT	not serious[1]	none	not serious[2]	none	none	⊕◯◯◯ **VERY LOW**	**[Seymour 2014]** Prehospital fluids did not reduce likelihood of ICU admission adjusted OR 0.64 (95% CI 0.37-1.10).
Impact of prehospital intravenous fluid therapy on mortality									
2	2313	non-RCT	not serious[1]	none	not serious[2]	very serious[3]	none	⊕◯◯◯ **VERY LOW**	**[Band 2011]** No significant difference in mortality was noted: adjusted RR 1.24 (95% CI 0.92 - 1.66) (p=0.16). **[Seymour 2014]** Prehospital fluids reduced hospital mortality adjusted OR 0.46 (95% CI 0.23-0.88) (p=0.02).

non-RCT: non-randomised controlled trial (observational study), IQR: inter quartile range, EMS: Emergency Medical Services, SD: standard deviation, MAP: mean arterial pressure, IV: intravascular, RR: risk ratio, 95% CI: 95% confidence interval' CVP: central venous pressure, EGDT: early goal directed therapy, OR: odds ratio, ICU: intensive care unit.

1. See risk of bias table
2. Single centre studies may limit generalizability.
3. Small study numbers limits precision/accuracy. Failure to report confidence intervals
4. Includes abstract only oublications. Publication bias likely (Guerra)
5. Abstract only publication, insufficient detail to rule out other bias

Impact of prehospital care on mortality

VERY LOW quality evidence was identified, from five non-randomised trials,[29 32 33 53 56] addressing mortality (see **table 3.10**). One of the studies was published in abstract form only.[29]

Band *et al*[53] did not find any significant difference in mortality adjusted RR 1.24 (95% CI 0.92 - 1.66, p=0.16). Femling *et al*[29] also failed to report any significant difference in their study comparing patients arriving via EMS with patients arriving by other means. They identified that 113/378 (30%, arrived by EMS) vs 34/107 (31%, arrived by other means) patients died, neither confidence intervals nor the p-value was given, but there was a statement that statistical analysis failed to achieve significance.

Guerra *et al*[32] noted that if EMS providers pre-alerted the hospital that a patient with sepsis was enroute to them that unadjusted mortality was improved OR 3.19 (95% CI 1.14– 8.88, p=0.04), however no adjusted point estimate was reported. McClelland *et al*[33] found a difference in mortality rates when EMS staff recognised and treated patients as septic compared to when sepsis was not recognised: 21% (EMS identified) versus 16% (not identified). Finally, Seymour *et al*[56] reported that prehospital vascular access reduced mortality adjusted OR 0.31 (95% CI 0.17 - 0.57, p<0.01).

Impact of prehospital antibiotic therapy

VERY LOW quality evidence was identified from one randomised controlled trial,[55] reporting the impact on outcomes for patients receiving prehospital antibiotics (see **table 3.10**). This study was published in abstract form only. Chamberlain[55] reported that prehospital antibiotic therapy leads to reduced ICU stay (mean ICU stay: 6.8±2.1 days (intervention) vs 11.2±5.2 days (control), p=0.001) and reduced mortality (28-day mortality: 42.4% (intervention) vs 56.7% (control); Odds Ratio (OR) 0.56; 95%CI 0.32-1.00).

Impact of prehospital intravascular fluid therapy

VERY LOW quality evidence was identified, from two non-randomised controlled trials,[53 56] reporting impact on outcomes for patients receiving prehospital fluid therapy (see **table 3.10**).

Band *et al*[53] reported that initiation of intravascular fluid therapy by EMS does not improve mortality (adjusted Risk Ratio (RR) 1.24; 95% CI 0.92 - 1.66). The

only study to demonstrate a positive impact following prehospital fluid administration among sepsis patients indicated that prehospital fluids were associated with reduced likelihood of organ failures (adjusted OR 0.58; 95%CI 0.34-0.98) and reduced hospital mortality (adjusted OR 0.46; 95%CI 0.23-0.88), but not reduced ICU admission (adjusted OR 0.64; 95%CI 0.37-1.10).[56]

3.5 Discussion

Limited evidence supports the argument that ambulance services may provide care for a "not insignificant" number of patients with sepsis. However, the evidence identified relates predominantly to the USA, with little robust evidence pertinent to the UK. International evidence indicates sepsis may account for more than 3% of ambulance workload. Furthermore, a small number of studies each report that patients admitted to Intensive Care from the Emergency Department are likely to have arrived at hospital via the ambulance service, suggesting ambulance services may care for a significant proportion of critically ill sepsis patients.

Identification of sepsis can be challenging. There is no single reliable test or marker to confirm sepsis. Potential predictors of sepsis have been summarised in **table 3.6**. The studies identified in this systematic review suggest accuracy of prehospital sepsis recognition by ambulance clinicians varies considerably. This variation could have numerous causes. In many areas paramedic education programs have not focussed sufficient attention on sepsis as a clinical syndrome and paramedic knowledge of sepsis is often poor.[17 28 38 112] It is possible that ambulance clinicians encounter sepsis patients earlier in their clinical course, before they become seriously ill, and it is also not known if in-hospital and prehospital clinical assessments, such as blood pressure, correlate in sepsis patients. An additional factor may be that routine in-hospital tests such as white cell count and lactate are not commonly used within EMS, which may limit the ability to extrapolate from in-hospital studies.

Many prehospital sepsis screening tools rely upon the Surviving Sepsis Campaign Systemic Inflammatory Response Syndrome (SIRS) criteria which were initially described to improve sepsis recognition in the ED and intensive care environments. Although SIRS describe physiologic signs marking the transition from infection to sepsis they lack specificity for sepsis. SIRS are observable following a wide variety of insults other than infection, leading some to question

the value of SIRS to identify sepsis.[16 20] Churpek *et al*[16] recently demonstrated that SIRS criteria were not reliable predictors of sepsis or mortality in the ward setting. Use of SIRS criteria to identify sepsis in the prehospital environment may therefore be equally ineffective.

The three studies documenting the development of prehospital screening tools for sepsis included more organ dysfunction criteria and included non-SIRS variables (see **table 3.6**). Among these tools sensitivity for severe sepsis ranged from 0.76 to 0.85 while specificity ranged from 0.47 to 0.86; they appear to perform better than tools based upon the SIRS diagnostic criteria. However, none have been clinically validated.

In-hospital data indicate that early identification and initiation of treatment can improve outcomes from severe sepsis. Recognition of sepsis in the patient's home or in the ambulance before arriving at the ED has the potential to further reduce time to diagnosis and potentially to reduce time to a limited number of treatments; this in turn could lead to further improved patient outcomes. There have been a limited number of studies that have attempted to demonstrate improved outcomes following prehospital recognition of sepsis.

Studnek *et al*[27] report that prehospital identification lead to a reduction in time to EGDT (69 vs 131 minutes, 95% CI not reported, p=0.001) as well as a reduction in time to antibiotic therapy (70 vs 122 minute, 95% CI not reported, p=0.003). Seymour *et al* (2010b) reported that if EMS administered intravenous fluids before arriving at hospital a larger proportion of severe sepsis patients achieved Mean Arterial Pressures >65mmHg within 6 hours (EMS IVF 70% vs no IVF 44%) however this result was not statistically significant (p=0.09). Neither of these studies demonstrated a reduction in mortality.

Bayer *et al*[102] estimated that median time to both intravenous fluid therapy and antibiotics was 19 minutes from call time when prehospital doctors took part in a small German study. However, this estimate assumed that both fluids and antibiotics would be administered within 10 minutes of arriving on scene and was not verified from records. Band *et al*[53] reported that if EMS personnel identified sepsis that antibiotics were administered 36 minutes sooner and fluids administered 34 minutes sooner following arrival at hospital.

The only study thus far that has demonstrated improved outcomes from prehospital intervention was conducted by Chamberlain[55] who showed if paramedics administered ceftriaxone in the ambulance that patients received this treatment 3.4±2.6 hours sooner than if they had waited until hospital arrival, and furthermore that 28-day mortality was reduced from 56.7% to 42.4% (OR 0.56, 95% CI 0.32 to 1.00, p=0.049).

One plausible explanation for the apparent failure to demonstrate improved outcomes could be that prehospital clinicians struggle to recognise sepsis and therefore do not intervene soon enough, aggressively enough, or frequently enough. Many prehospital sepsis studies identified in this systematic review rely on an adaptation of traditional in-hospital diagnostic criteria. However, the prehospital environment presents unique challenges which limit the ability to extrapolate from studies conducted in hospitals. For example, the time window for assessment is limited, patients present earlier in their course of illness and there is limited access to supportive investigations such as white blood cell count, microscopy/culture or diagnostic imaging.

3.6 Limitations

This systematic review employed a broad search strategy to capture as much published literature as possible. Inclusion criteria were intentionally not restrictive to include as much of the evidence base as possible. Despite using very broad search criteria, little robust evidence regarding sepsis in the prehospital environment was identified. The studies found employed disparate methodologies, exhibit significant heterogeneity, generally involve small numbers of patients (limiting the precision of reported results) and were invariably of very low quality. The conclusions that can be drawn from this systematic review are therefore limited and findings should be interpreted with caution.

3.7 Recently published evidence

This systematic review summarises the findings of searches for evidence published before July 2015. The chapter was not formally updated prior to submission of the thesis in November 2017 as two publications (see **Appendix 1.1** and **1.2**) had already been secured in 2016. This section serves to update the

chapter with more recent publications relevant to the topic, without formally updating the chapter.

Two further papers reporting the incidence of sepsis among ambulance transports were identified. Tusgul *et al* identified that among 11411 patients transported to a university hospital, 3.8% were confirmed to have sepsis in the ED.[113] Intheir study, Walchok *et al*[114] reported that 1185 of 56643 (2.1%) of patients transported by EMS received a Sepsis alert. Of these, 848 received an admitting diagnosis of sepsis (1.5%). However, the authors do not report the proportion of missed cases, that is the proportion of patients who received a sepsis diagnosis at hospital but did not receive a sepsis alert from EMS.

One study that reported the correlation of individual variables with the outcome of sepsis. Hunter *et al*[115] reported that the area under the ROC curve predicting sepsis dependent upon end tidal carbon dioxide ($EtCO_2$) was 0.99 (95% CI 0.99-1.00) and for temperature was 0.64 (95% CI 0.64-0.71). The area under the ROC curve predicting severe sepsis for $EtCO_2$ was 0.80 (95% CI 0.73-0.86), for temperature was 0.41 (95% CI 0.33-0.49), for SBP was 0.65 (95% CI 0.57-0.73), for DBP was 0.64 (95% CI 0.55-0.72), and for $Sp0_2$ was 0.59 (95% CI 0.52-0.68). [115]

Two studies addressed the use of lactate and $EtCO_2$ to improve detection of sepsis. In an observational study of 112 patients meeting SIRS criteria for sepsis (heart rate >90bpm, Respiratory rate >20 bpm, body temperature <36˚C or >38˚C), Boland *et al*[116] assessed the potential beneficial effect of obtaining lactate readings during prehospital assessments. Elevated prehospital lactate was not associated with diagnosis of sepsis, increased hospital admission, ICU admission, length of hospital stay or mortality. They concluded that including lactate measurements did not achieve a level of diagnostic accuracy in identifying patients with severe sepsis or septic shock that would warrant measurement of lactate in the field. [116]

In a methodologically similar study, Hunter *et al*[115] assessed the potential beneficial effect of including $EtCO_2$ readings, to improve sepsis stratification by paramedics. Their study included 330 patients meeting sepsis alert criteria of heart rate >90bpm, Respiratory rate >20 bpm and body temperature <36˚C or >38˚C. Patients were subsequently stratified according to $EtC0_2$. 147 patients

had $EtC0_2$>25mmHg, while 183 patients had $EtC0_2$=<25mmHg. For patients meeting the sepsis alert criteria with EtCO2=<25mmHg, sensitivity for predicting sepsis was 0.69 (95% CI 0.62-0.75), the specificity was 0.67 (95% CI 0.57-0.75), the PPV was 0.78 (95% CI 0.70-0.84), and the NPV was 0.99 (95% CI 0.92-1.0). The sensitivity for predicting severe sepsis was 0.90 (95% CI 081-0.95), the specificity was 0.58 (95% CI 0.52-0.65), the PPV was 0.47 (95% CI 0.39-0.97), and the NPV was 0.93 (95% CI 0.87-0.97). The sensitivity for predicting mortality was 0.76 (95% CI 0.54-0.90), the specificity was 0.46 (95% CI 0.40-0.52), the PPV was 0.11 (95% CI 0.07-0.18), and the NPV was 0.95% (95% CI 0.90-0.98). The sensitivity for predicting ICU admission was 0.67 (95% CI 0.57-0.76), the specificity was 0.50 (95% CI 0.43-0.57), the PPV was 0.41 (95% CI 0.33-0.49), and the NPV was 0.75 (95% CI 0.66-0. 82). [115]

Three recent studies report the effectiveness of screening tools. Dorsett *et al*[117] reported a series of ED 1255 patients, of whom 560 did not arrive by EMS. Following exclusions, only 152 patients were included. Of these, 71 had infection, 38 had sepsis and 43 had severe sepsis or septic shock. Among these patients, performance characteristics for the qSOFA score were as follows, sensitivity 0.16 (95% CI 0.07–0.31), specificity 0.97 (95% CI 0.92–0.99), positive likelihood ratio (PLR) 5.91 (95% CI 1.6–21.8) and negative likelihood ratio (NLR) 0.86 (95% CI 0.8 – 1.0). [117] The authors conclude that qSOFA has poor sensitivity to detect sepsis in the prehospital environment.[117]

Jouffrey *et al*[118] reported a very small series of 37 patients with septic shock attended by Paris SAMU. SAMU is a prehospital system that includes a doctor as part of the clinical team. Nine patients died by day 28. Performance characteristics of different screening tools were as follows (95% CI not reported): [118]

Modified Robson Tool: sensitivity 100%, specificity 16%, PPV 39%, NPV 100%
qSOFA>=2: sensitivity 62%, specificity 16%, PPV 29%, NPV 44%
MEWS>=5: sensitivity 85%, specificity 33%, PPV 41%, NPV 80%
PRESEP>=4: sensitivity 92%, specificity 29%, PPV 41%, NPV 88%

Tusgul *et al* reported that a pre-hospital qSOFA>=2 had a sensitivity of 36.3% for ICU admission, 17.4% for ICU stay of three days or more and 68.0% for 48-hour mortality.[113] The sensitivity of two or more SIRS criteria reached 68.8% for ICU

admission, 74.6% for ICU stay of three days or more and 64.0% for 48-hour mortality. Confidence intervals were not reported.[113]

Three additional papers address the impact of prehospital care on subsequent management in hospital. Axelsson *et al*[119] reported on a cohort of 696 patients who attended the ED who were subsequently diagnosed with bacteraemia. They noted 308 patients arrived by EMS, infection was suspected in 32% of these cases but only 6% of cases were identified as possible septicaemia by the EMS nurse. Patients arriving by ambulance received antibiotics over 2 hours sooner (3 hours 15 min vs 58 min). Impact on outcomes was not reported.[119]

Carberry *et al*[120] noted that the introduction of a sepsis screening tool, in the Scottish Ambulance Service in the Lanarkshire region, reduced the time to triage by 82% (17 minutes vs 3 minutes) and reduced the time to antibiotics by 39% (49 minutes vs 30 minutes). They failed to report impact on outcomes.[120]

Results of the PHANTASi study were published online in late Novemebr 2017. [121] This study enrolled 2698 patients, with 1535 in the intervention arm and 1137 in the usual care arm. Patients in the intervention arm received prehospital antibiotics while those in the usual care arm received antibiotics at hospital. The primary outcome was all cause mortality at 28 days. The intervention group received antibiotics a median of 26 minutes before arriving at hospital whereas the usual care group received antibiotics 70 minutes after arriving at hospital. There was no difference in outcomes at 28 days. In the intervention group, 120 (8%) of patients died. In the usual care group 93 (8%) died (relative risk 0·95, 95% CI 0·74–1·24). [121]

These additional publications are consistent with the findings of the systematic review. An additional systematic review of prehospital recognition and management of sepsis was published by Lane *et al*[122] that also confirms the findings reported previously. The signal from the various papers is consistent with the findings reported in the publications arising from this chapter.

3.8 Conclusion

This chapter has reported a large systematic review. It identifies that there is little robust evidence addressing sepsis in the prehospital environment. The available evidence is generally of very low quality, and more often than not,

comprises very small study populations. The review has collated available data concerning the incidence of sepsis in the prehospital environment. Known predictor variables have been identified and diagnostic accuracy with respect to recognition of sepsis by ambulance clinicians has been reported, while accuracy of existing prehospital sepsis screening tools has also been established.

Sepsis patients represent a small, but important, group of patients managed by ambulance services, potentially up to 3% of workload. Ambulance clinicians could play an important role in reducing time to identification of sepsis for this group of patients. However, recognition of sepsis in the prehospital environment is challenging. At present the available data suggest that ambulance clinicians struggle to identify patients with sepsis, and that ambulance clinicians would benefit from further education to improve their understanding of sepsis.

This systematic review further demonstrates that early identification of sepsis in the prehospital environment can lead to improvements in processes of care, and expediting of patient passage through the emergency care pathway. Thus far, however, the evidence indicates that early recognition alone does not translate into improved clinical outcomes for patients. Research indicating that prehospital antibiotic therapy and fluid resuscitation leads to improved patient outcomes is currently lacking. Research quantifying the impact of prehospital antibiotic administration on subsequent hospital care, and appropriate antibiotic stewardship is also lacking. There may be scope for ambulance clinicians to deliver a limited number of important interventions, such as intravenous fluid therapy and antimicrobial therapy, provided it can be demonstrated that ambulance clinicians are able to consistently recognise patients with sepsis.

The development and validation of a reliable prehospital sepsis screening tool has been called for and should be considered a priority for improving prehospital care of sepsis patients. Thereafter, well conducted studies addressing diagnostic accuracy and key clinical interventions, such as antibiotic administration and fluid resuscitation will be required.

Chapter 4

Study data

4.1 Introduction

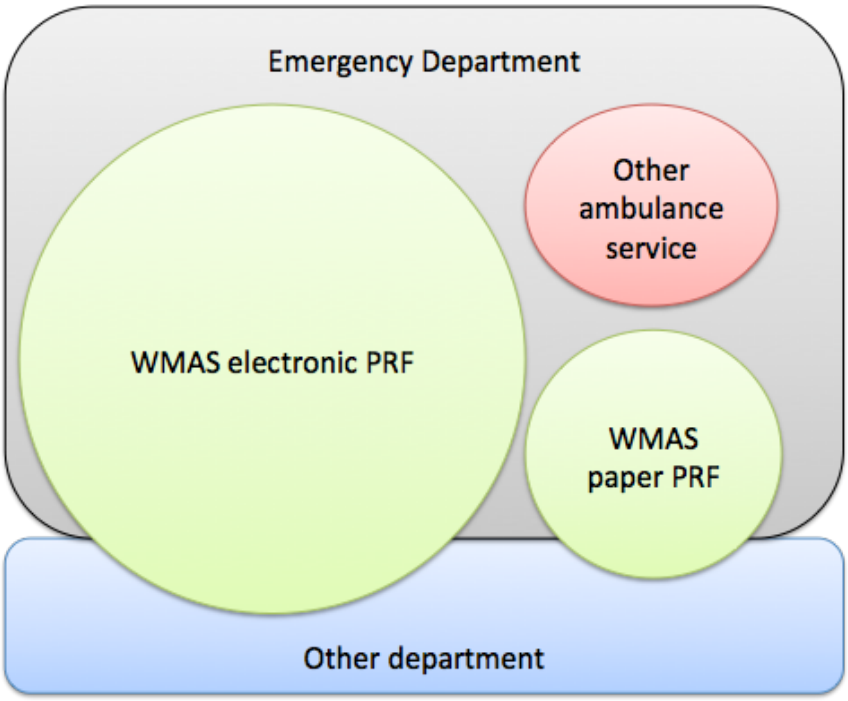
This chapter describes the datasets used in this study. It will also address how the data were cleaned and processed to ensure subsequent analyses were reliable.

4.2 Datasets

This study utilises two distinct datasets. The first dataset was supplied by West Midlands Ambulance Service NHS Foundation Trust (WMAS) and comprises all patients managed by WMAS and transported to University Hospital North Staffordshire NHS Foundation Trust (UHNS) between 01 July 2013 and 30 June 2014. The second dataset is supplied by UHNS and comprises all patients attending the Emergency Department (ED) who arrived by ambulance between 01 July 2013 and 01 July 2014. An additional day of ED data was requested to ensure that ambulance cases assigned just prior to midnight on 30 June 2014, but arriving at hospital following midnight on 01 July 2014 could be appropriately paired. The study period comprises 01 July 2013 to 30 June 2014. Each patient visit is considered a unique case.

Most cases supplied by WMAS were transported to the ED at UHNS. However, a proportion were transported to other departments, for example the Surgical Assessment Unit. Similarly, most cases supplied by UHNS ED were patients who arrived by WMAS ambulance, but a small proportion will have arrived by a different ambulance provider, for example St John Ambulance (see **figure 4.1**). The overwhelming majority of ambulance cases arriving at the ED on 01 July 2014 will have no corresponding ambulance record, as the call to the ambulance service is likely to originate on 01 July 2014 and these data were not provided.

Figure 4.1 Relationship between datasets



PRF – patient report form

4.2.1 Ambulance Service data

During the study period, two distinct methods for documenting patient care were in operation within WMAS. The primary method of documenting patient care, accounting for approximately 75% of patient encounters in the region, was the electronic patient report form (ePRF). Electronic Care Solution (ECS) was based upon Siren ePCR®, developed by Medusa Medical Technologies of Halifax, Canada. The remaining cases were documented on paper patient report forms (pPRF's). Per WMAS operational instruction,[123] pPRF's are only to be used when the responding ambulance resource is not fitted with the required tablet computer, or there is a failure of the tablet computer.

Data from the ePRF are stored in a relational database housed on the NHS Spine. The NHS Spine supports the Information Technology (IT) infrastructure for the NHS for Health and Social Care and is maintained by NHS Digital. It allows information to be shared securely through national services such as the Electonic Prescription Service, Summary Care Record and the e-Referral Service. Paper PRF's are collated at local ambulance station level and transported, securely, to WMAS headquarters where they are scanned. Specialist scanning software is used to extract data from the pPRF and import them into a Microsoft Excel®

spreadsheet. There is approximately eight weeks' delay between completion of the clinical case and electronic availability of data from pPRFs.

4.2.1.1 Data requested from the Ambulance Service

All WMAS records for patients transported to UHNS between 01 July 2013 and 30 June 2014 were requested. Detail concerning which data fields were requested, and which were provided, are reported in **table 4.1**.

Table 4.1 Data requested from WMAS

Category	Variable	Provided
Incident details	Date of call	Yes
	Time of call	Yes
	Ambulance incident number	Yes
	Location address (including postcode)	Yes
	Dispatch Complaint	Yes
	Dispatch Priority	Yes
Ambulance details	Call sign	Yes
	Time assigned	Yes
	Time mobile	Yes
	Time arrived	Yes
	Time leaving scene	Yes
	Time at hospital	Yes
	Hospital department	No
Patient details	Name	Yes
	Surname	Yes
	Age	Yes
	Date of birth	Yes
	Gender	Yes
	Ethnicity	Yes[1]
	Address (including post code)	Yes
	Past medical history	No
Clinical observations	Respiratory rate	Yes
	Peripheral oxygen saturations (SpO_2)	Yes
	End tidal carbon dioxide ($EtCO_2$)	Yes[1]
	Heart rate	Yes
	Blood pressure	Yes
	Capillary bed refill time (CBRT)	Yes
	Blood glucose	Yes
	Temperature	Yes
	Glasgow coma score (GCS)	Yes
	Alert/Verbal/Pain/Unresponsive score (AVPU)	Yes
	Pupil response	Yes
	Pupil size	Yes
Interventions	Vascular access	Yes
	Endotracheal intubation	Yes
	Mechanical ventilation	No
Treatments	Oxygen therapy	Yes
	Fluid therapy	Yes
Outcome	Crew diagnosis	Yes[2]

1 – very little data provided, very high proportion of missing data
2 – only generic physiological category e.g. cardiac provided

4.2.1.2 Missing Ambulance Service data

All the missing data fields were visible within the online individual ePRF record viewer, however WMAS IT were unable to locate which fields within the database were being used to store the data, to facilitate extraction. I was assured that WMAS IT would identify and extract the missing data fields upon implementation of a new patient record database in June 2016 (following introduction of a new ePRF). Consequently, I did not resort to manual review of the records to update the supplied ambulance data with the missing data fields. In July 2016 I was informed that WMAS would not be able to extract the missing data fields as this work had not yet been scheduled for completion by the new ePRF vendor. Unfortunately, the 'old' database, on the NHS spine, had been decommissioned and was no longer available to allow manual review. As of August 2017, these data are still unavailable.

4.2.1.3 Source of data provided by the Ambulance Service

After a several months delay, WMAS confirmed they were unable to provide any data from pPRF's. This was due to an error in the configuration of the Optical Character Recognition (OCR) software used by WMAS that had not previously been identified. Ambulance crews document transport destination by entering a unique hospital code in a specified field of the pPRF. In the case of UHNS the code is "UNS". Unfortunately, the OCR software used to extract the data from the pPRF had not been configured correctly. It had been configured to read the three *letters* as three *integers* and consequently the OCR software failed to recognise any hospital codes. The study approvals specify that data will be provided for patients transported to UHNS only. As it was not possible for WMAS to limit the extraction of data to the subset of patients transported UHNS, it was not possible to obtain data from patients whose care was documented on a pPRF.

Data from the ePRF were extracted, however there were several issues with the data provided. As previously noted, the database of patient records is housed on the NHS Spine, and is not maintained by WMAS. Furthermore, the database used is a commercial product designed primarily for hospital use. As such, there are several hundred fields that are not appropriate to the ambulance record (for example results of hospital laboratory tests or diagnostic imaging), and many fields required by ambulance services were not defined within the commercial product (for example ambulance arrival at scene time). Several unused fields within the commercial database were re-mapped to enable recording of key

ambulance data. However, no map of where data was stored was available to the WMAS Information Technology (IT) staff. Consequently, WMAS IT staff struggled to locate and extract all the data fields requested.

Clinical records for 45292 patients were provided in several Microsoft Excel® files. WMAS reported that extraction of the requested data over the NHS Spine was complex and time consuming. As a result, data were extracted one month at a time. Over a 12-week period 12 Microsoft Excel® files were provided, each of which related to one month of the year for which the data were requested. The 12 files were subsequently merged by Michael Smyth (MAS) into a single Microsoft Excel® file that was used for data cleaning and data linkage.

4.2.2 Emergency Department data

Documentation of care in the Emergency Department (ED) at UHNS is via the Emergency Department Information System (EDIS) PatientFirst® software. EDIS PatientFirst® is an electronic health record that facilitates a complete electronic record of care provided in the Emergency Department. However, in practice, it was not implemented as a paperless system at UHNS. Patient details and initial triage are all recorded electronically. Thereafter, a paper copy is printed and the ED doctor/clinician annotates this paper copy with the medical history, pertinent findings and treatments delivered. These hand-written notes are then scanned and appended as an image file to the initial electronic record.

4.2.2.1 Data requested from the Emergency Department

All ED records for patients who arrived by ambulance at UHNS between 01 July 2013 and 01 July 2014 were requested. Detail concerning which data fields were requested, and which were provided, are reported in **table 4.2**. Clinical records for 48534 patients were supplied as a Microsoft Excel® file.

4.2.2.2 Missing Emergency Department data

Informal discussion with UHNS revealed that much of the missing data was stored within the electronic patient record, but not in an accessible format. As previously mentioned, clinicians frequently enter hand written notes or append scanned images into the patient's electronic record. These additions cannot be searched electronically. Unfortunately, resources to manually extract this information and append it to the electronic record were not available.

Table 4.2 Data requested from UHNS

Category	Variable	Provided
Hospital arrival	Date	Yes
	Time	Yes
	Hospital record number	Yes
	Ambulance call sign	No
Patient details	Name	Yes
	Surname	Yes
	Age	Yes
	Date of birth	Yes
	Gender	Yes
	Ethnicity	Yes
	Address (including post code)	Yes
Clinical observations	Respiratory rate	Yes
	Peripheral oxygen saturations (SpO_2)	Yes
	End tidal carbon dioxide ($EtCO_2$)	No
	Heart rate	Yes
	Blood pressure	Yes
	Capillary bed refill time (CBRT)	No
	Temperature	Yes
	Blood glucose	Yes
	Glasgow Coma Score (GCS)	Yes[1]
	Alert/verbal/pain/unresponsive score (AVPU)	No
	Pupil reaction	No
	Pupil Size	Yes
	Partial pressure of oxygen (PaO_2)(in ED)	No
	Partial pressure of carbon dioxide ($PaCO_2$)(in ED)	No
	Lactate (in ED)	No
	Creatinine (in ED)	No
Laboratory observations (where available)	White blood cell count (WCC)	Yes
	Platelet count	Yes
	International Normalised Ratio (INR)	Yes
	Prothrombin time (PTT)	Yes
	Creatinine	Yes
	Lactate	Yes[2]
	Base deficit	No
	C-Reactive Protein (CRP)	Yes
	Causative pathogen	No
Interventions	Peripheral vascular access	No
	Central venous access	No
	Endotracheal intubation	No
	Mechanical ventilation	No
Treatments	Oxygen therapy	No
	Fluid therapy	No
	Antibiotic therapy	No
	Vasoactive drug administration	No
Outcome	ED diagnosis	Yes
	Discharged or admitted?	Yes
	Admission details	Yes
	Survived or died	Yes
	Date of death	Yes
	Time of death	Yes
	Location of death	Yes

1 – high proportion missing,
2 - very low proportion reported

Consequently, the availability of data was dependent upon how information was entered into the record. For example, when the hospital laboratory service were reporting results of laboratory analysis, appropriate data fields within the electronic record were remotely updated. However, when blood analysis was undertaken by the clinician in the ED, pertinent results were hand written by the attending clinician in the narrative section, or the print out from the blood gas analysis machine is scanned as an image file, rather than individual values being entered in the appropriate data field of the electronic patient record (as occurs with results from the laboratory).

4.3 Data cleaning

Data cleaning is a term without a well-established definition. Although data cleaning targets errors in data, the definition of what is, and what is not an error, can be highly subjective, depending upon the application, and purpose of the intended analysis.[124] Rahm *et al*[125] describe data cleaning as the process of detecting and removing errors and inconsistencies from data, in order to improve data quality. Chu *et al*[126] assert that analysis of clean data facilitates improved decision making, whereas analysis of 'dirty' data can generate erroneous understandings and incorrect decisions. Chu *et al*[126] propose that data cleaning is comprised of two stages: error detection and error repair. Muller and Freytag[124] suggest that data errors be classified as errors in syntax, semantics or coverage.

4.3.1 Errors in the data

4.3.1.1 Errors of syntax

Lexical anomalies refer to discrepancies between the data content and the expected format. For example, one would expect an integer value in a variable reporting age, however a case may have an entry recorded as "male" suggesting the data has been erroneously inputted in to that field. Such anomalies may be amenable to correction, depending on the structure of the remaining variables.

Domain format anomalies refer to discrepancies in the format of a variable. For example, a name variable might be formatted as "John Smith", but the entry appears as "Smith, John". Such data entries are not incorrect, but require cleaning to ensure formatting is consistent across all cases.

Data irregularities refer to the non-uniform use of values, units and abbreviations. For example, reporting a pressure in kilopascals (kPa) rather than millimetres of mercury (mmHg). These data are not incorrect, but will require conversion to the appropriate units.

4.3.1.2 Errors of semantics

Integrity constraint anomalies refer to entries that violate a 'rule' in relation to a variable. For example, a negative age value violates the expectation that age will have a minimum of 0 (in years). *Integrity constraints* enhance our understanding of the data by restricting the data to a set of valid instances. Each constraint is a rule representing knowledge about the domain and the values allowed for representing certain facts.

Contradictions are violations of a relationship between the variables for a particular case. For example, a discrepancy between the variables Age and Year of Birth. Contradictions are either violations of functional dependencies that can be represented as *integrity constraints* or *duplicates* with inexact values.

Duplicates are two or more observations representing the same entity within the dataset. The values of these observations need not be identical. Inexact *duplicates* are instances of *contradiction* between two or more observations. They represent the same entity but with differing values for all, or some, of its variables. This hardens the detection of duplicates and their mergence.

Invalid entries are by far the most challenging anomalies found within any dataset. These errors cannot be categorised as one of the above anomalies. However, such entries still do not represent valid entities from the dataset. Invalid entries result from our inability to describe reality within a formal model using *integrity constraints*. They are extremely hard to find, and even more complicated to correct. There are no rules (or constraints) which are violated by these entries and we only have incomplete knowledge about every entity in the dataset.

4.3.1.3 Errors of coverage

Missing values occur when individual cases have missing data, commonly attributable from omissions during the initial data collection or because of errors occurring during data extraction phase.

Missing cases occur when all data pertaining to a patient is missing from the dataset.

4.3.2 Errors in WMAS data

The primary WMAS dataset includes 45292 observations comprising 50 variables (see **table 4.3**).

Table 4.3 Original structure of WMAS data

Variable name	Type	Variable name	Type
PatientID	character	Heart Rate (BPM)	integer
Forename	character	Systolic BP (mmHg)	integer
Surname	character	Diastolic BP (mmHg)	integer
Address 1	character	Temperature (Celcius)	number
Address 2	character	Pain Score	integer
City	character	BloodGlucose	number
County	character	Glasgow Coma Scale	integer
Post Code	character	GCS Eye	character
DOB	date	GCS Verbal	character
Sex	character	GCS Motor	character
Incident Address	character	Skin Colour	character
Incident City	character	Cap Refill Time	integer
Incident Postcode	character	AVPU	character
Incident Time	time	Pupil Right Reactivity	character
Response Priority	character	Pupil Right Size	integer
Mobile	time	Pupil Left Reactivity	character
Arrive Scene	time	Pupil Left Size	integer
Leaving Scene	time	Airway Adjunct	character
Destination Arrival	time	Airway Signs	character
Chief Complaint	character	Airway Status	character
Secondary Complaint	character	Arrest Occurance	character
Respiratory Rate (BPM)	integer	Impression	character
Oxygen Saturation (%)	integer	Crew Level 1	character
Peak Flow (l/min)	integer	Crew Level 2	character
End Tidal CO2 (kPa)	number	Institution Name	character

A second, supplementary table addressing treatments administered was provided, comprising 44302 observations of 35 variables (see **table 4.4**).

Table 4.4 Drugs administered

Variable	Variable
Patient ID	Hydrocortisone
Adrenaline 1:1000	Ibuprofen
Adrenaline 1:10000	Ipratropium
Amiodarone	Lidocaine
Aspirin	Low Molecular Weight Heparin
Atropine	Metoclopramide
Benzyl Penicillin	Morphine Sulphate
Buccal suscard	Naloxone
Chlorphenamine	Oramorph
Dextrose 40% oral gel	Oxygen
Diazemuls	Paracetamol
Diazepam	Paracetamol suspension
Entonox	Salbutamol
Furosemide	Sodium Chloride 0.9%
Glucagon	Sodium Chloride 5ml amps
GlucoGel	Tenecteplase
Glucose 10%	Water for Injection
GTN spray	

4.3.2.1 Errors of syntax

A small proportion of *Lexical anomalies* were identified. Inspection of the WMAS data revealed a sample of observations where the temperature and blood glucose readings appeared to be transposed. That is, blood glucose appeared to be documented in the temperature field and vice versa. Manual review of the individual ePRF cases via the online viewer enabled review of narrative and diagnostic fields to confirm entries were transposed. For example, where blood glucose was reported to be 37.9 mmol/L and temperature 5.8°C and there was no mention in the narrative of hyperglycaemia or past medical history of diabetes, nor any mention of hypothermia with documentation of normal skin colour that was warm to touch. Such cases were considered transposition errors. In addition, a substantial proportion of these cases had repeat observations with readings in the normal range in the appropriate fields. These errors are assumed to be due to user input as the two fields are next to one another on the ePRF. No *domain format anomalies* were identified.

A small number of *data irregularities* were identified. Inspection of the `WMAS` data revealed a sample of `temperature` readings that were much higher than expected. It would be very unusual indeed to identify a patient with a temperature more than 42°C. However, a group of readings incompatible with life were found, ranging between 96° and 102° degrees. Manual review of these records via the online ePRF viewer revealed these readings had been incorrectly documented in Fahrenheit rather than Celsius. Affected entries were converted to celsius by applying the formula $T_{(°C)} = \left(T_{(°F)} - 32\right) * \frac{5}{9}$ and rounded to one decimal place. It is unclear why the ePRF entry was in Fahrenheit.

4.3.2.2 Errors of semantics

Numerous *integrity constraint anomalies* were identified. All numeric variables, apart from `peak flow` and `end tidal CO2` (due to very high missingness), were assessed to determine if they had observations that were outside "normal" ranges (see **table 4.5**). Where a potential anomaly was identified, the case was manually reviewed via the online ePRF viewer. In many cases, the out of expected range reading was confirmed to be correct. Reviewing subsequent values and the narrative often confirmed that the "suspect" reading was correct. For example, if systolic blood pressure was above 180mmHg (outside the "normal" range), but the past medical history and narrative mentioned hypertension, and there were subsequent blood pressure readings outside the "normal" range to corroborate the queried reading, then the "suspect" reading was, in fact, deemed correct. In others, it was obvious that there was an error. For example, a respiratory rate of 122, and within the narrative section of the ePRF a normal respiratory pattern was documented, with subsequent respiratory rate readings in the range of 12 - 16. In many cases these errors are likely to be due to typing error, such as pressing the "2" key twice rather than once.

Ambulance crews are expected to document at least two sets of observations,[44] [123] consequently where such errors were identified, if a second set of observations was available, the suspect reading was replaced with the value from the next available set of observations. If no second set of observations was available the suspect reading was deleted.

Table 4.5 Integrity constraint thresholds WMAS

Variable name	Manual review threshold
Date of birth	<01-01-1914 (equates to age>100)
Respiratory Rate (BPM)	>60
Oxygen Saturation (%)	<80 or >100%
Heart Rate (BPM)	>180
Systolic BP (mmHg)	>160 or <60
Diastolic BP (mmHg)	>120 or < 40
Temperature (Celsius)	<35 or >41
Pain Score	>10
Blood Glucose	<1 or >12
Glasgow Coma Scale	<3 or >15
Cap Refill Time	>5
Pupil Right Size	>10
Pupil Left Size	>10

A small number of *contradictions* were identified. All AVPU readings were compared with components of the Glasgow Coma Score. All cases with a score of "alert" or "no response" on AVPU were compared with the Glasgow Coma Scale variable. All cases with a score of "verbal response" on AVPU were compared with the GCS verbal variable. All cases with a score of "pain response" on AVPU were compared with the GCS motor variable. Where discrepancies were identified, the case was manually reviewed via the online ePRF viewer. It should be noted that it is possible for discrepancies to occur between these two variables as they are often recorded at different time points. Consequently, only obvious errors were corrected such as AVPU equal to "conscious" with Glasgow Coma Scale below eight and a narrative indicating the patient was unresponsive on initial assessment.

Within the WMAS dataset it was possible to identify several thousand *duplicate cases*. *Duplicate cases* commonly arise when ambulance crews fail to finalise cases on the ePRF in a timely manner. Usual practice is for the first clinician on scene, for example the Rapid Response Vehicle (RRV) paramedic, to initiate treatment and to begin documenting care provided on the ePRF. Following arrival of the ambulance crew, the RRV paramedic would hand over continuing

care to the ambulance crew, then finalise and close their ePRF record. The transporting ambulance crew would continue treatment whilst transporting the patient to hospital. To document care provided the transporting ambulance crew re-open the case generated by the RRV paramedic and continue in the same ePRF record. However, if the RRV paramedic fails to finalise and close his or her ePRF before the ambulance crew commence documenting their care in the ePRF, a new ePRF case, with a different identification number, is generated. In addition, the RRV paramedics entry will not be imported into the ambulance crew record. As such, the patient will now have two ePRF records in the database, one initiated by the RRV paramedic and another by the ambulance crew, relating to the same incident. Such cases are usually easy to identify as the incident time, incident location details and patient details are the same. However, inspection of both ePRF records is required to confirm that the ePRF entries relate to the same patient, and ensure that the apparent duplicate ePRF records do not relate to multiple patients at the same incident, as might occur following a road traffic collision.

Sorting the WMAS dataset by date and time of incident facilitates rapid identification of potential duplicate cases. Sorting was performed within Microsoft Excel® utilising "sort and filter' function with the "expand selection" option to ensure row integrity was maintained. All potential duplicate cases were manually reviewed and where appropriate, the records were merged to ensure the maximum amount of data were captured. The clinical record of the first on scene, that is the record with the earliest time for `arrived scene`, was considered the primary record and any additional information from the second or subsequent records was appended to the first. For example, it is not uncommon for the first on scene to obtain the patient's name and age, but not the date of birth details, as they are not essential to initiate treatment. These are commonly completed in the ambulance record when more time is available. So, the first record was updated with information from subsequent records where appropriate. The first recorded clinical observations, such as pulse, blood pressure and blood glucose reading, were used in all cases. After duplicate cases had been harmonised the superfluous records were deleted. Following the manual review of cases, 6809 duplicate cases had been removed and 38483 unique patient cases remained.

No *invalid entries* were identified. All numeric variables, apart from `peak flow` and `end tidal CO2` (due to very high missingness), were assessed to determine if they had observations that were negative numbers (see **table 4.5**), in addition all categorical clinical variables were assessed to determine if they had observations that were outside of "permitted" ranges (see **table 4.6**).

Table 4.6 Valid categorical entries

Variable	Valid entries
`GCS_eye`	Spontaneous, responds to voice, responds to pain, no response
`GCS_verbal`	Oriented, confused, inappropriate words, incomprehensible sounds, no response
`GCS_motor`	Obeys, localizes, withdraws, flexion, extension, no response
`Skin`	Normal, cyanosed, flushed, pallor, mottled, rash, jaundiced
`CBRT`	Normal, delayed
`AVPU`	Alert, responds to voice, responds to pain, no response
`RPupilReact`	Brisk, fixed, sluggish
`LPupilReact`	Brisk, fixed, sluggish
`Airway Adjunct`	OPA, NPA, LMA, ETI
`Airway Signs`	Patent, partial obstruction, obstructed

4.3.2.3 Errors of coverage

Thousands of cases were identified with *missing values.* Handling of missing data within the WMAS dataset is addressed in Chapter 6 Missing Data.

It was not possible to confirm how many patient records (*missing cases)* there were. The WMAS dataset was ordered chronologically and reviewed to ensure cases existed for each consecutive day. No days were identified where there were no cases.

4.3.3 Errors in UHNS data

The UHNS dataset includes 48 534 observations comprising 43 variables (see **table 4.7**).

Table 4.7 Original structure of UHNS data

Variable	Variable
ID_Number	DischargeDate
AttendanceDate	DischargeTime
AttendanceTime	Disposal
Category	Pulse
FirstName	Resps
Surname	SpO2
Address1	SBP
Address2	DBP
PostCode	GCS-eye
Sex	GCS-verbal
DoB	GCS-motor
Age	Temp
Outcome	BM
Date_Discharged	CBRT
Time_Discharged_ED	WCC
Diagnosis1	CRP wide
Diagnosis2	platelet count
Diagnosis3	lactate
Diagnosis4	INR
Diagnosis5	Bilirubin
Did_Patient_Die_In_ED	Creatinine
Did_Patient_Die_In_Hospital	

4.3.3.1 Errors of syntax

No *lexical anomalies* were identified. No *domain format anomalies* were identified. No *data irregularities* were identified.

4.3.3.2 Errors of semantics

Numerous *integrity constraint anomalies* were identified. All numeric variables were assessed to determine if they had observations that were outside of a

"normal" range (see **table 4.8**). Where a potential anomaly was identified, the full observation was reviewed to determine if the suspect entry was consistent with the other variables. Where it was, obvious there was an error, for example, a respiratory rate of 120, with other observations like pulse oximetry, heart rate and blood pressure all in the normal range, the entry was deleted. It was not possible to review the records electronically or to request a third party to do so.

Table 4.8 Integrity constraint thresholds UHNS

Variable name	Potential error threshold
Date of birth	<01-01-1914 (equates to age>100)
Pulse	>220
Resps	>80
SpO2	<70 or >100
SBP	>200
DBP	>140
Temp	<35 or >41
BM	>10
CBRT	>5 seconds
WCC	>20000
CRP wide	>200
platelet count	<10
lactate	>10
INR	>5
Bilirubin	>300
Creatinine	>600

No *contradictions* were identified. Each patient is assigned a unique identification number on arrival at the ED. No *duplicate* cases were identified. No *data irregularities* were identified.

No *invalid entries* were identified. All numeric variables were assessed to determine if they had observations that were negative numbers. In addition, all categorical clinical variables were assessed to determine if they had observations that were outside of a "permitted" range (see **table 4.9**).

Table 4.9 Valid entries

Variable	Valid entries
GCS Eye	Spontaneous, voice, pain, no response
GCS Verbal	Oriented, confused, inappropriate words, incomprehensible sounds, no response
GCS Motor	Obeys, localises, withdraws, decorticate, decerebrate, no response

4.3.3.3 Errors of coverage

Only a very small proportion of cases had data for all variables. Most cases had one or more missing data points. This occurred because many variables were optional. For example, although there were five diagnostic variables, it was only required to have one diagnosis with the remaining four being empty. Similarly, laboratory tests are only required for a subset of patients. The very high missingness of data meant imputation was not practical, as imputation in such circumstances may not yield reliable results.[127]

The UHNS dataset is required to establish the hospital diagnosis. Missing data were presumed to be clinically normal when modelling the UHNS data for clinical diagnosis. In other words, if any data point required for classification was missing, the missing value was assumed to be in the normal, healthy range. For example, when calculating a National Early Warning (NEWS) score, a score is assigned depending on the oxygen saturations. Peripheral oxygen saturations are attributed a score of 3, 2, 1 or 0 dependinare categorised as being below 92% or in the ranges 92% to 93%, 94% to 95% or above 95%, and respectively. Where a NEWS score was being calculated for a patient with a missing oxygen saturation value, the missing value would be assumed to be in the normal range (above 95%) and score a '0' was applied. This approach is equivalent to single value imputation with normal value substitution, an approach that has previously been employed among critically ill patients.[107 128] Normal value substitution may lead to an underestimation of the true prevalence of sepsis in the population being studied.

It was not possible to confirm how many, if any, patient cases were missing from the dataset provided *(missing cases)*. The UHNS dataset was ordered chronologically and reviewed to ensure cases existed for each consecutive day. No days were identified where there were no cases.

4.4 Calculated fields

4.4.1 WMAS calculated fields

Several additional fields were 'calculated' and appended to the WMAS dataset (see **table 4.10**):

Table 4.10 Calculated fields (WMAS)

Variable	How calculated
`Oxygen`	Calculation of the NEWS score includes use of supplemental oxygen. These data were drawn from the supplemental treatments table
`Fluids`	Calculation of the SOFA score includes the need for fluid resuscitation. These data were drawn from the supplemental treatments table
`Age`	Required to determine if age is correlated with sepsis. Determined as difference (in years) between `Incident Time` and `Date of Birth`. Stored as integer (rounded down)
`Location`	Required to determine if place of residence is correlated with sepsis. Determined from the `Incident PostCode`, address `PostCode`

4.4.2 UHNS calculated fields

Several additional fields were 'calculated' and appended to the UHNS dataset (see **table 4.11**):

Table 4.11 Calculated Fields (UHNS)

Variable	Purpose
ED_Infection	Presence of infection is required to diagnose sepsis. Determined from ED_Diagnosis1, ED_Diagnosis2, ED_Diagnosis3, ED_Diagnosis4, and ED_Diagnosis5.
DrSepsis	Variable used to store sepsis diagnoses by the ED doctor. Determined from ED_Diagnosis1, ED_Diagnosis2, ED_Diagnosis3, ED_Diagnosis4, and ED_Diagnosis5. Required as potential reference standard.
GCS_sum	GCS is required to calculate several sepsis related scores score.
MAP	MAP is required to calculate SOFA score.
NEWS	The NEWS score is required to determine UK Sepsis Trust 2016 sepsis status.
UHNS_SIRS	The number of SIRS criteria are required to determine UK Sepsis Trust 2014 sepsis status.
UHNS_SIRS_Sepsis	Required as potential reference standard.
UHNS_NICE_Sepsis_Risk	Required as reference standard.
UHNS_Organ_Failure_Score	The number of organ failures are required to determine organ failure sepsis status.
UHNS_Organ_Failure_Sepsis	Required as potential reference standard.
UHNS_qSOFA_Score	The number of qSOFA score required to determine qSOFA sepsis status.
UHNS_qSOFA_Sepsis	Required as potential reference standard.
UHNS_SOFA_Score	The number of SOFA score required to determine SOFA sepsis status.
UHNS_SOFA_Sepsis	Required as potential reference standard.

4.4 Conclusion

This chapter has described the nature of the datasets as received from WMAS and UHNS. It has also detailed how the data were processed and cleaned to prepare them for record linkage, which is described in the next chapter.

Chapter 5

Record linkage

5.1 Introduction

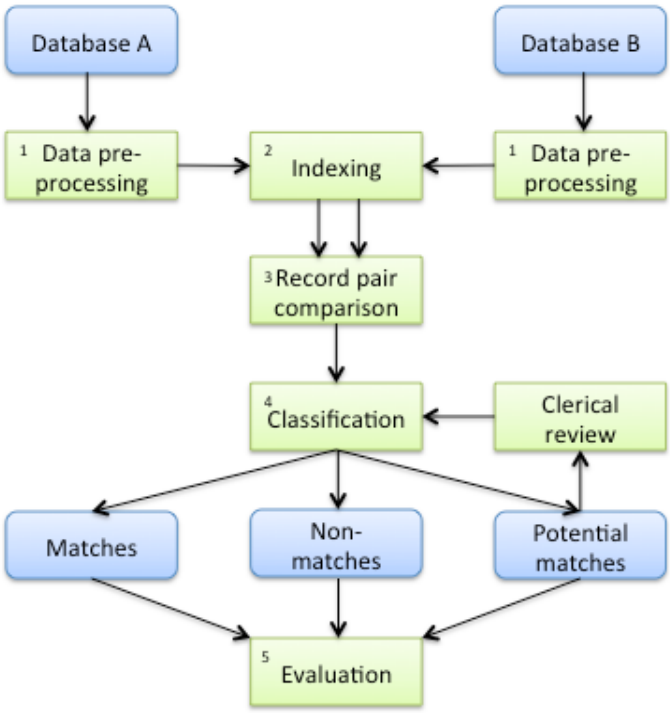
This study utilises patient records relating to a single health care episode to develop a prehospital sepsis screening tool. Data pertinent to the health care episode are spread between two datasets, one from the Ambulance Service and the other from the Emergency Department. The process of pairing a patients' ambulance record with their corresponding hospital record is referred to as record linkage. Record linkage is defined as the process of bringing together records, relating to the same entity, recorded in different locations, usually from different databases.[129] The purpose of record linkage is to create richer, more detailed records that can be used in subsequent analysis.[130] This chapter will provide an overview of the record linkage process, describe how study datasets were were linked in this study and evaluate the success of the linkage exercise.

Christen[129] described the record linking process as comprising five key steps as per **figure 5.1.**

- During *data pre-processing* the format of data fields within each of the databases are standardised to ensure compatibility.
- *Indexing* minimises the number of records being compared, to facilitate the efficient and effective generation of potential record pairs.
- During *record pair comparison,* key variables are evaluated using a variety of comparison functions.
- During *classification* candidate record pairs are classified as *matches*, *non-matches* or *potential matches* the latter of which mandates clerical review.
- During *evaluation* the complexity of the record linking exercise is determined, and the quality and completeness of the matched data are assessed.[129]

Several statistical software packages, such as Stata® and R®, have specialist functions or packages to perform record linkage. In addition, bespoke record linkage software packages, such as Link Plus® (as used in this study), may be employed.

Figure 5.1 Data linkage process



reproduced from Christen (2012)

In the context of this study, record linkage brings together patient records from a single Ambulance Service and a single Hospital Emergency Department. Records relate to a specific health care episode, initiated when a patient called '999'. The Ambulance Service responded, initiated treatment and conveyed the patient to the Hospital Emergency Department, where further investigations were undertaken and treatments continued. Record linkage brings together data concerning the ambulance and hospital episode for use in development of a prehospital sepsis screening tool, detailed in chapters 7 and 8.

5.2 Background

5.2.1 Data pre-processing

Data within different databases can be stored in differing formats, and may require standardisation before being used for record linkage. The objective of *data pre-processing* is to ensure that the data being compared across the two databases have a consistent structure and format, to optimise the comparison

process. Common examples of data variations that may adversely impact record linkage are shown in **table 5.1**.

Table 5.1 Common record field standardisations

Field	Concern	Examples
Names	Case	John Smith vs JOHN SMITH
	Nicknames	Charles vs Chuck
	Synonyms	William vs Bill
	Prefixes	Dr John Smith
	Suffixes	John Smith, Esq.
	Punctuation	O'Malley, Smith-Taylor
	Initials	JR Smith vs J.R. Smith vs John Robert
	Transposition	John Smith vs Smith, John
Address	Abbreviations	Rd vs Road
Dates	Format	01-01-2017 vs 01JAN2017
	Invalid values	day=32, month=13, year=2025
Identification numbers	Format	99999999 vs 99-999-999 vs 99 999 999
Geographic detail	Abbreviation	UK vs United Kingdom
Gender	Format	male/Female vs M/F vs 1/2

Fields used for record linkage invariably utilise personal information including names and addresses that can originate from several potential sources. The information may be dictated by a patient, while a data entry clerk inputs the information; it may be transcribed from an audio recording, where the data entry clerk is not able to clarify spellings; it may be entered from scanned notes, where the handwriting may not be clear (or indeed it may be incorrectly spelt) or it may be output from optical character recognition (OCR) software. Such data are prone to error, and these fields require appropriate cleaning and standardisation before record linkage is attempted.

Christen[129] identifies that there can be up to four stages in data pre-processing:

- *Removal of unwanted characters and words* entails removal of superfluous punctuation characters including commas, quotation marks and excess white space. Extraction of data from a database into a different format may also introduce additional characters or 'stop words' that require removal. For example, special characters, such as letters with an umlaut are encoded using unicode numbers (that is ë, Unicode number U+00EA), that may not be understood by all software packages resulting in incorrect insertions, or superfluous characters. For example, Microsoft Excel® commonly inserts a question mark before uncertain data entries.
- *Expand abbreviations and correct spellings* to reduce variation, for example 'Rd' becomes 'Road', in addition obvious spelling mistakes should be corrected.
- *Segment attributes into well-defined and consistent output attributes* It can be beneficial to parse fields with several important sub-fields into multiple fields. Most commonly this occurs with an *address* field where the single data field is divided into *house number*, *street*, *suburb*, *city*, *postcode*, and *county* fields. Parsing fields into distinct units allows the linker to maximise the available information and achieve partial agreement when record pairs do not agree character for character. This can be particularly important to account for changes across time, such as name changes after marriage or address changes after moving. In such cases, matching on the separate pieces allows for the possibility of partial credit which, when combined with other information, may provide sufficient evidence that the records being compared relate to the same person.[131]
- *Verification of correctness of attribute values* where a reliable external data source is available, both databases being linked can be checked against the reliable data. For example, the Royal Mail produces a database of postcode addresses that relate a postcode, to a street address. These data could be used to verify and check the addresses as recorded in both databases, and it may be possible to amend incorrect entries before matching is attempted.

5.2.2 Indexing

When linking records in two databases, the potential number of comparison pairs grows quadratically with the number of records in the databases to be matched. In other words, two databases containing *m* and *n* records respectively, will generate *m*n* comparison pairs. In contrast, the number of possible true matches increases linearly. In other words, the maximum number of true matches is *m* or *n,* whichever is the smaller of the two.[129] This is because every record from database one should be compared with every record in database two, yet only a small number of records from database two should match any single record from database one - the overwhelming majority of comparison pairs will, in fact, be *non-matches*. This has significant implications for the computational power required to link records.

If two databases with one million records each were to be linked, this would result in one trillion (1000000*1000000) comparison pairs. Even if 100000 comparisons could be performed per second it would still take almost 116 days to compare the two databases.[129] In order to reduce this inefficiency, *indexing* is used to significantly reduce the likelihood of *non-matches*. This minimises the computational overheads and the time taken to compare records. *Indexing* filters out record pairs that are unlikely to be *matches*, and generates candidate record pairs that will be examined in greater detail during the *comparison* stage.

The most common approach to *indexing* is *blocking*. *Blocking* reduces the number of candidate pairs through focusing on specified agreement patterns.[132] Without *blocking*, record one, from database one, would be compared to every record in database two. However, with *blocking*, record one, from database one, is only compared with a subset of records from database two. *Blocking* is achieved by defining a *blocking variable* that is used to filter records in the second database.

A common choice for *blocking variable* is *surname*. Thus, if record one, from database one, has a *surname* of 'Miller', then the only records used for comparison will be those records from database two, with the surname 'Miller'. In practice, if an exact match were to be used for *blocking* then the potential for missed matches would be very high. Consequently, exact matching is seldom used. Rather, phonetic encoding algorithms, such as Soundex or New York State Identification and Intelligence System (NYSIIS) are commonly employed when *blocking* with a surname field.

Phonetic encoding maps words from natural language to strings representing their pronunciation (the phonetic code). Words which sound similar generate the same phonetic code.[132] As a result, record one, of database one (surname = "Miller"), will be compared with all records of database two where the surname sounds like "Miller". That is, record one from database one will be compared with records from database two with surnames such as "Millar", "Millor", "Millie" etc. An additional benefit of phonetic encoding is that it reduces the likelihood of missed matches due to minor misspellings.[130] Because obvious non-matches are not compared, *blocking* significantly reducing the computational overhead and time taken for processing.[129]

5.2.3 Record pair comparison

Even if considerable effort has been expended during the *data pre-processing* stage, it is highly likely that records corresponding to true matches, that is relating to the same entity, can contain different attribute values across the two different databases. A common example would be that of a person who married, and changed their surname, with pre-and post-martial surnames being recorded in each of two databases. If the only field used for comparison is *surname* then all such true matches would most likely be missed. It is vital therefore that data matching algorithms employ comparison functions that return some indication of how similar two records are overall.

Candidate pairs identified during *indexing* require further assessment over several additional fields to determine potential match status. Additional pre-specified *matching variables* enable comparison of records to determine wider similarity between candidate pairs. Common *matching variables* include *first names*, *date of birth*, *postcodes* and *unique personal identification numbers* such as NHS number. The greater the similarity across the *matching variables*, the more likely it is that the records correspond to the same entity. Techniques have been developed to account for minor misspellings and typographical errors within *matching variables*. The most important of these are:

- Strings can be converted to phonetic codes (Soundex or New York State Immunisation Information System) before comparison.
- Strings can be compared using editing distance techniques to determine how many steps (insertions, deletions, transpositions, etc.) are required to get from String A to String B. For example, it

would require one step--one character deletion--to get from "Billy" to "Bill".

- Names can be linked to an array of synonyms ("William" and "Bill") to account for the use of nicknames.

The above methods ensure minor variations within fields do not result in an automatic *non-match* status being assigned when the records are formally compared. There are two distinct approaches to formal record comparison: deterministic and probabilistic.

Deterministic algorithms provide a binary outcome when comparing a given set of *matching variables*, the algorithm may be implemented as a single-step or multi-step strategy. In a single-step strategy, the *matching variables* of candidate pairs are compared all at once. If the *matching variables* of the candidate pair agree, character for character, on all *matching variables* and the candidate pair is uniquely identified (no other record pair matched on the same set of values), then the record pair is classified as a *match*. A candidate pair is classified as a *non-match* if any of the *matching variables* disagree, or if the candidate pair is not uniquely identified. Because deterministic linkage requires exact agreement on the specified *matching variables*, it can have a very high rate of missed-matches (false negatives), as any recording errors or missing values within the *matching variables* will result in a *non-match* pair. However, the rate of false-matches (false positives) can be very low indeed as unrelated records are unlikely to have exact matches on the *matching variables* by chance alone.[133]

To reduce the number of missed matches, a multi-step deterministic strategy may be employed. In a multiple-step strategy, also referred to as an iterative or stepwise strategy, candidate pairs are matched in a series of progressively less restrictive steps in which candidate pairs that do not meet a first round of matching criteria are passed to a second round of matching criteria for further comparison. If a candidate pair meets the criteria in any step, it is classified as a *match*. Otherwise, it is classified as a *non-match.*[131]

In addition to the potentially high false negative rate, the deterministic approach fails to consider that certain *matching variables* may have greater discriminatory power than others. For example, matching unique personal identifiers, such as

NHS number, will provide greater confidence of a true *match* than would a match based on date of birth.

Probabilistic strategies, first proposed by Fellegi and Sunter,[134] assess the discriminatory power of each *matching variable* in addition to *matching* the content of the *matching variables*. The weighting of each variable depends on the discriminatory power of the variable, so that agreements on important variables (for example the NHS Number is intended to be unique) makes a larger contribution than agreement on lesser variables (for example the probability that date of birth matches by chance is 1 in 365).[135]

The *matching variables* of each candidate pair identified during the *indexing* phase are compared, producing an agreement pattern. The weight assigned to agreement or disagreement on each *matching variable* is assessed as a likelihood ratio, comparing the probability that true matches agree on the *matching variable* ("*m*-probability"), to the probability that false matches randomly agree on the *matching variable* ("*u*-probability").

Where the candidate pair agree on a *matching variable*, an agreement weight is calculated by dividing the *m*-probability by the *u*-probability and taking the log to base 2 ($\log_2$) of the quotient. For example, if the probability that true matches agree on month of birth is 97%, and the probability that false matches randomly agree on month of birth is 8.3% (1/12), then the agreement weight for month of birth would be calculated as ($\log_2$[0.97/0.083]), which equates to 3.54. When the candidate pair disagree on a *matching variable*, a disagreement weight is calculated by dividing 1 minus the *m*-probability by 1 minus the *u*-probability. For example, the disagreement weight for month of birth would be calculated as ($\log_2$ [(1-0.97)/(1-0.083)]), or - 4.93.[131]

The initial work of Fellegi and Sunter[134] has been extended. Porter and Winkler demonstrated that it is possible to modify the *m*- and *u*-probabilities to reflect similarities of the *matching variables,* between 0.0 and 1.0, calculated by approximate string comparison algorithms. Their results showed significant improvements in matching quality rather than reliance on a binary *match* or *non-match* dependent upon the exact string matches as advocated by Fellegi and Sunter.[136 137]

A second extension to the probabalistic approach focusses upon frequently occurring values within *matching variables*. The rationale underpinning this idea asserts that the more frequent an attribute value is in a database, the less discriminative this value is for classifying a candidate pair as a *match* or *non-match*. For example, the surname value 'Smith' is likely to occur much more frequently than 'Dijkstra' in the West Midlands region of the United Kingdom. Consequently, a candidate pair with the surname 'Dijkstra' is more likely to be a match than a candidate pair with the surname 'Smith'.

By generating comparison scores (likelihood ratios) for the *blocking key* and each of the *matching variables*, we can generate an overall record comparison score for the two records being compared. The example tables below illustrate how these record comparison scores might be calculated.

It can be clearly seen that candidate pairs 1 and 2 relate to the same and different persons respectively. Candidate pairs 3 and 4 however present a much greater challenge despite very similar overall scores. Candidate pair 3 could potentially be a sibling or married couple pairing whereas candidate pair 4 most likely does relate to a true match. The calculated record comparison scores above can be used to assign provisional *match*, *non-match* and *potential match* status for candidate pairs for scores within certain specified ranges. For example, if the score is greater than four provisionally assign *match* status, if the score is below two provisionally assign *non-match* status and if the score is between two and four provisionally assign *potential match* status. Assessment of these provisional allocations is detailed in the next section. Note - scores attributed to variables being compared have been arbitrarily assigned for illustrative purposes ranging from 0 (no match) to 1.0 (perfect match).

Table 5.2 Candidate pair 1

Record	Name	Surname	DoB	Gender	Post code	
A1	James	Lee	17/02/70	Male	B31 2RP	
B777	James	Lee	17/02/70	Male	B31 2RP	***Total***
Score	*1.0*	*1.0*	*1.0*	*1.0*	*1.0*	***5.0***

Table 5.3 Candidate pair 2

Record	Name	Surname	DoB	Gender	Post code	
A1	James	Lee	17/02/70	Male	B31 2RP	
B666	David	Leigh	30/10/83	Male	B27 5QX	***Total***
Score	*0*	*0.4*	*0*	*1.0*	*0.1*	***1.5***

Table 5.4 Candidate pair 3

Record	Name	Surname	DoB	Gender	Post code	
A1	James	Lee	17/02/70	Male	B31 2RP	
B555	J	Lee	07/03/72	Female	B31 2RP	***Total***
Score	*0.1*	*1.0*	*0.3*	*0*	*1.0*	***3.3***

Table 5.5 Candidate pair 4

Record	Name	Surname	DoB	Gender	Post code	
A1	James	Lee	17/02/70	Male	B31 2RP	
A1	Jim	Lee		Male	B31 2RP	***Total***
Score	*0.6*	*1.0*	*0*	*1.0*	*1.0*	***3.6***

5.2.4 Record pair classification

Provisional record pair classification generates three potential outcomes of *match*, *non-match* or *potential match*. *Matches* are assumed to refer to the same entity, *non-matches* refer to two different entities and *potential matches* are those candidate pairs where the match status is not clear, and which require further *clerical review* to determine the final match status. All records not paired during *indexing* are implicitly classified as *non-matches*. Following record comparison, all provisional record pairs are re-evaluated and classified as either *match* or *non-match*. This can be by manual review, or automated by defining thresholds, above and below which, record pairs are automatically assigned *match* or *non-match* status. All *potential matches* require manual review to re-classify them as either *match* or *non-match.*

In a perfect linkage, all paired records are correctly classified as *matches*, and all non-paired records are classified as *non-matches*. If record pairs are misclassified, error is introduced. False matches occur when records from different entities are incorrectly assigned *match* status. Missed matches occur when records from the same entity fail to link and are assigned *non-match* status.

In statistical parlance, a *match* is the equivalent of a 'True Positive', a *non-match* the equivalent of a 'True Negative', a *false-match* the equivalent of a 'False Positive' and a *missed-match* the equivalent of a 'False Negative'.

5.2.5 Evaluation of matched data

Following clerical review, when all candidate pairs have been formally classified into *matches* and *non-matches* the results of matching should ideally be assessed for both *matching quality* and *matching completeness*. *Matching quality* refers to the proportion of identified matches that are *true matches*, or in statistical parlance, minimising the proportion of false positives. *Matching completeness* is concerned with the proportion of *true matches* within the databases that were correctly paired, or minimising the proportion false negatives/missed cases.

All the preceding steps, *data pre-processing*, *indexing*, *comparison* and *classification* all have an impact on *matching quality* and *matching completeness*. *Data pre-processing* ensures consistent data format to enable comparison, *indexing* filters out records that are unlikely to be a match, while detailed *comparison* ensures similarity between records can be quantifiably assessed. The *comparison* and *classification* stages clearly have a major impact upon *quality*; in particular, *clerical review* can be very challenging when the linker is required to make a manual *match* or *non-match* decision when multiple *matching variables* have values that differ from each other. *Indexing* will have a major impact on *completeness* due to filtering out cases not to be matched - filtering ensures records will not be paired and will result in *non-match* classifications being assigned without records being compared.

To reliably report *quality* and *completeness*, the true status of matching should be determined. However, in practice, it may be exceedingly difficult to establish the true status unless some external mechanism exists to confirm the findings, e.g. contacting individuals to verify the data. As such these measures of linkage effectiveness are commonly missing.

Accuracy is not recommended when reporting how successful record linkage has been. *Accuracy* is calculated from the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN):

$$accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

In data matching, the number of *non-match* (TN) cases far exceeds the number of *match* (TP) cases, which will significantly bias reported accuracy. For example, suppose two databases each containing 500000 records are to be matched, of which 250000 are true matches. We run our record linkage program *indexing* generates say, 12500000 candidate pairs, and our user classifies 300000 of these as *matches* and 12200000 as *non-matches*. We will assume we are lucky and have a gold standard data to confirm the true status of our matches are can report that of the 300000 *matches*, 200000 are true positives (TP), 100000 are false positives (FP), furthermore 50000 of the *non-matches* were incorrectly classified and should have been *matches* - these are our false negatives (FN). The remaining 11850000 *non-match* pairs are true negatives (TN). Accuracy calculated for this example is 96%, despite identifying only 200000 of a potential 250000 true positive cases.

$$\begin{aligned} accuracy &= \frac{(TP + TN)}{(TP + FP + TN + FN)} \\ &= \frac{(12050000)}{(12500000)} \\ &= 0.964 \end{aligned}$$

More appropriate measures to report are *precision*, *recall* and *F-measure*.

Precision, the proportion of true positives, calculated for this example is 67%.

$$\begin{aligned} precision &= \frac{TP}{(TP + FP)} \\ &= \frac{200000}{300000} \\ &= 0.67 \end{aligned}$$

Recall, the proportion of true matches identified as true matches, calculated for this example is 80%.

$$\begin{aligned} recall &= \frac{TP}{(TP+FN)} \\ &= \frac{200000}{250000} \\ &= 0.80 \end{aligned}$$

F-measure is a balance of *recall* and *recall*. For this example, the *F-measure* is 73% and is calculated as follows:

$$\begin{aligned} F-measure &= 2*\frac{(precision*recall)}{(precision+recall)} \\ &= 2*\frac{(0.67*0.80)}{(0.67+0.80)} \\ &= 0.729 \end{aligned}$$

Finally, the complexity of a data matching exercise may also be reported. *Complexity* is defined as the number of candidate pairs identified (c) divided by the product of number of records in database A (m) and number of records in database B (n). *Complexity* provides a measure of how effective indexing has been in reducing the number of *non-matches*. Lower complexity values indicate more successful record linkage. *Complexity* calculated in our hypothetical example is 0.005%, suggesting indexing has been very successful.

$$\begin{aligned} complexity &= \frac{c}{(m*m)} \\ &= \frac{12500000}{500000*500000} \\ &= 0.00005 \end{aligned}$$

5.3 Methods

5.3.1 Pre-processing of WMAS data

Fields from ambulance service data, used during record linkage, and all pre-processing performed are outlined in **table 5.6** below. Patient identifiable data including patients name, surname, date of birth and home address post code were required to link personal records. Ambulance incident date was required to ensure correct instances were linked together i.e. to ensure ambulance and hospital records for patient who attended hospital on multiple occasions

throughout the 12-month period were linked appropriately. Suffixes and Prefixes were not removed as Link Plus removes these automatically.

Table 5.6 Preprocessing of ambulance data

Variable name	Definition	Variable type	Pre-processing
Amb_Forename	text	Patients forename	Convert to lower case
			Remove all punctuation
			Remove excess white space
Amb_Surname	text	Patients surname	Convert to lower case
			Remove all punctuation
			Remove excess white space
Amb_DOB	date	Patients date of birth	Remove excess white space
			Convert to DD-MM-YYYY format
			Delete invalid values
Amb_Gender	text	Patients gender	Remove all white space
			Convert to male or female
Amb_Pt_Postcode	text	Patients post code	Convert to lower case
			Remove all punctuation
			Remove all white space
Amb_Incident_Date	date	Date of 999 call	Remove all white space
			Convert to DD-MM-YYYY format
			Delete invalid values

5.3.2 Pre-processing of UHNS data

Fields from emergency department data, used during record linkage, and all pre-processing performed are outlined in **table 5.7** below:

Table 5.7 Pre-processing of Emergency Department data

Variable name	Definition	Variable type	Pre-processing
UHNS_Forename	text	Patients forename	Convert to lower case
			Remove all punctuation
			Remove excess white space
UHNS_Surname	text	Patients surname	Convert to lower case
			Remove all punctuation
			Remove excess white space
UHNS_DOB	date	Patients date of birth	Remove excess white space
			Convert to DD-MM-YYYY format
			Delete invalid values
UHNS_Gender	text	Patients gender	Remove all white space
			Convert to male or female
UHNS_Pt_Postcode	text	Patients post code	Convert to lower case
			Remove all punctuation
			Remove all white space
UHNS_Incident_Date	date	Date ED attended	Remove all white space
			Convert to DD-MM-YYYY format
			Delete invalid values

Patient identifiable data including patients name, surname, date of birth and home address post code were required to link personal records. Date of emergency department attendance was required to ensure correct instances were linked together. Suffixes and Prefixes were not removed as Link Plus removes these automatically.

5.3.3 Link Plus Software

Link Plus is a Windows-based software program developed by the Cancer Division of the Centres for Disease Control and Prevention (CDC) in the United States. It is a probabalistic linkage program designed to match records in registries. Although originally designed for use with cancer registries, there is no barrier to using Link Plus with data other than cancer registry data.

In 2004, the CDC released the first production version of Link Plus, Version 1.0, which probabilistically linked files, and allowed 1-1 matching, but provided no support for manual review of comparison pairs. The second production version of Link Plus, Version 2.0, was released in 2007 and allowed 1-1 matching and included powerful support for the manual review process. The third version of Link Plus is only available in beta-version and has the following enhancements:

- It removes the limitation on the number of records included in File 2; the program works for any number of records in File 2 if the computer has sufficient memory to read in data from File 1 (estimated limitation of 4.5-4.8 million records for the file size of File 1).
- Users can choose whether to write all potential matches (many-to-many linkages) or only the matches with the highest score to the linkage report (1-to-many linkages).
- It provides confirmation-like matching method for variables such as address where an exact match contributes a positive weight to the total linkage score while no match contributes nothing (zero weight) to the total score.
- It provides Social Security Number-like matching method for a generic ID which allows partial matching by calculating similarity.
- It provides refined name matching methods that are more robust against the frequency of names or outlier of names, and can handle the names with embedded spaces and apply partial matching algorithms to swapped names.

Link Plus version 3 (beta) was used to link records in this study.

5.3.4 Choice of blocking variable

Surname was chosen to be the *blocking variable*. Link Plus has two options for phonetic encoding - Soundex and New York State Identification and Intelligence System (NYSIIS). NYSIIS was chosen for phonetic encoding as it has a reported 2.7% increase in accuracy over Soundex.[138]

5.3.5 Choice of matching variables

Variables chosen to be *matching variables* are reported in **table 5.8**.

Table 5.8 Variables used as *matching variables* during record linkage

Ambulance Service	Hospital	Method
WMAS_Forename	UHNS_Forename	NYSIIS
WMAS_DOB	UHNS_DOB	Date
WMAS_Gender	UHNS_Gender	Exact
WMAS_Postcode	UHNS_Postcode	String
WMAS_Incident_Date	UHNS_Incident_Date	Date

5.4 Results

5.4.1 Generation of candidate pairs

Link Plus software generated 35382 candidate pairs. No linkage score threshold was defined to assign a preliminary *match* or *non-match* status. All candidate record pairs were subject to manual review.

5.4.2 Clerical review of candidate pairs

The ambulance incident date and hospital arrival date are not generated by user input, they are system generated. As such, they are less prone to error. Consequently, all candidate pairs were required to have identical incident dates. The only exception to this rule would occur when the '999' call was received before midnight, but the ambulance arrived at hospital shortly after midnight. Where a difference of one day was identified, the ambulance and hospital records were manually reviewed to determine if the times were consistent with this mechanism. Any candidate pair not meeting this requirement was categorised as a *non-match*.

It is highly unlikely that two individuals with almost identical personal details would have used the same ambulance service and attended the same Emergency Department on the same day. Generally, any one of date of birth, gender or postcode could contain an error, or be missing, provided there was agreement on forename, surname and incident date. Minor misspellings between forename and

surname fields were tolerated provided there was agreement on the remaining fields. Cases where forename or surname were missing, or clearly different, were categorised as *non-match*.

Following clerical review 33289 candidate pairs were categorised as *matches,* while 2093 were classified as *false matches*. Several thousand records were unmatched, 5194 ambulance records had no corresponding hospital record, while no ambulance record could be identified for 15428 emergency department records.

5.4.3 Evaluation of record linkage

No gold standard data is available to determine which, if any, *match* records were false positives (FP) and which *non-match* records were false negatives (FN). It is therefore not possible to report *precision*, *recall* or *F-measure* for this record linking exercise.

Complexity for indexing is as follows:

$$complexity = \frac{c}{(m * m)}$$
$$= \frac{35382}{(38483) * (48534)}$$
$$= 0.00000189$$

Of 38483 ambulance records, 33289 (86.5%) were successfully linked with an emergency department record. Manual review of 250 randomly selected record pairs failed to identify any incorrect linkages. Results of this linkage exercise compare favourably with similar EMS linkage studies. Newgard[139] undertook a probabilistic linkage of EMS and trauma registry data with a 95.9% success rate. Downing *et al*[140] linked ambulance and Emergency Department (ED) records relating to assault in the United Kingdom (UK), reporting 84.2% successful linkage. Mears *et al*[141] linked Emergency Medical Services data with a state-wide stroke registry and reported 63% of cases correctly linked. Finally, to develop their Critical Illness Score, Seymour *et al*[107] report successfully linking 85% of cases, spanning a ten year period, using the same Link Plus software as used in this study.

Of the 5194 (13.5%) unlinked ambulance cases within this study, a significant proportion are likely to be patients who were transported to hospital

destinations other than the Emergency Department (ED). Manual inspection of 120 unlinked cases confirmed that 97 (80.8%) were transported to hospital destinations other than the Emergency Department. Further work to extract the hospital destination was promised by WMAS IT, however these data never materialised. As reported in **Chapter 4**, the ability to review WMAS records online is no longer available to enable manual review the remaining unlinked cases, thus it has not been possible to determine if the same proportion of records, in the remaining 5074 unreviewed cases, were transported to non-ED destinations. However, if the proportion were found to be similar this would suggest that 4328 of the unlinked cases could have been transported to destinations other than the ED, while around 866 would be unlinked ED cases. Such an adjustment would improve successful data linkage to 33289 of 34155 cases (97.5%), similar to that reported by Newgard.[139]

A small proportion of unlinked cases are likely to result from an ambulance crew being unable to identify a patient in the early stages of a health care episode e.g. when the patient has a significantly altered level of consciousness and is unable to confirm their name or address. In such cases the ambulance crew generally enter the name as 'unknown' or leave the relevant fields blank. At hospital however, the record can be updated once the patient has recovered sufficiently enough to provide these details, or has been identified through other means e.g. relatives or police. Consequently, a mismatch between the ambulance and hospital records is to be expected in a small proportion of cases. There were 58 such instances within the 5194 records (1.1% of unlinked cases) where the name and surname fields of the ambulance record were "missing" or "unknown".

5.5 Removal of identifiable data

Permission to process patient identifiable information without consent was obtained from the Health Research Authority (HRA) Confidentiality Advisory Group (CAG) under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 on 19 May 2014 (CAG reference: CAG4-03(PR2)2014). The conditions of approval specified that personal identifiable information would only be held for 6 months and that once ambulance and hospital records were linked all identifiable data would be removed. Record Linkage generated a unique record linkage key that was inserted into both datasets to enable future pairing of linked records. Following completion of the

record linkage exercise the following fields were removed from the respective datasets:

5.5.1 Ambulance data

- Incident details comprising all incident address fields including postcode.
- Patient details comprising forename, surname, date of birth, all address fields including postcode.

5.5.2 Emergency Department data

- Patient details comprising forename, surname, date of birth, all address fields including postcode.

5.6 Conclusion

This chapter has provided an overview of the record linkage process. It has detailed the pre-processing of data prior to record linkage, identified which variables were selected as blocking and matching variables as well as the approach to variable comparison. Comparison pairs were evaluated and the effectiveness of the linkage exercise was explored as far as practicable given the available data. The next chapter will focus upon management of missing data to prior to statistical analysis being performed.

Chapter 6

Missing data

"The only really good solution to the missing data problem is not to have any" (Paul D. Allison, 2001)

6.1 Introduction

Missing data are ubiquitous in clinical research.[142] Historically, analyses of datasets with missing values has commonly proceed with an assumption, either explicit or implicit, that the mechanism causing the missing data could be ignored.[143] However, such assumptions can significantly bias any findings. For example, if study participants with poorer clinical outcomes are more likely to drop out of a study before completion, then subsequent statistical analysis will be biased if it fails to adequately account for the poor outcomes of those participants who dropped out. Furthermore, missing data also has a significant impact on which statistical analyses can be undertaken as many statistical techniques presume complete data for all variables.[144]

Where data are incomplete, datasets usually require processing, to ensure that statistical analysis is undertaken with complete cases.[144] This chapter addresses missingness of data within the datasets used, provides an overview of the strengths and weaknesses of the different options for management of missing data, and describes how missing data were handled within this study.

6.2 Background

6.2.1 Identification of missing data

In addition to identification of cells with missing data, appropriate management of missing data entails determining if there is a justification for, or pattern to, the missingness of data. For example, a survey question asks if a participant has ever smoked marijuana. Where the response is "No", the participant is directed to skip subsequent questions addressing recent use, and proceed to the next section. Many researchers would not consider the skipped questions to be missing, as never having smoked marijuana implies no recent use. However, there will be no recorded values in the dataset for the skipped questions - how

are these "missing" values to be handled when performing analyses that rely upon complete data.[145] Consequently, although identification of missing values is straight forward, categorising the cause of missing data can be a challenge, and is required to ensure appropriate statistical methods are employed.

6.2.2 Categorising the cause of missing data

Rubin's[143] seminal work addressing missing data defined three categories thereof:

- ***Missing completely at random (MCAR):*** MCAR indicates data are missing independent of both observed and unobserved (missing) data. For example, blood test observations missing due to accidental damage to test tubes - the cause of missingness is not related to any variable in the dataset, but the random occurrence of accidental damage. The missing data are in effect a random subset of the dataset.
- ***Missing at random (MAR)****:* MAR indicates that missing data are related to the observed data but the pattern of missingness cannot be predicted. For example, a study is undertaken that requires two measurements of a variable be taken at the same time. The study protocol dictates that if the two measurements differ by more than a specified amount, a third measurement must be taken. A third measurement will thus be required in all cases where the difference between the two measurements is greater than the specified amount, but missing in those cases where the difference is less than the specified amount. Missingness of measurement three is therefore dependent upon the observed data (measurements one and two), however the occurrence of missing values cannot be predicted.
- ***Missing not at random (MNAR):*** MNAR indicates missing data are related to missing data values themselves. For example, clinicians omit an expensive diagnostic test as they can make a diagnosis without the additional information provided by the test.

If the data are MCAR or MAR: the missing data mechanism is said to be ignorable, this implies there is no need to model the missing data mechanism when generating replacement values.

If the data are MNAR: the missing data mechanism is said to be non-ignorable, this implies the missing data mechanism must be modelled when generating replacement values, and requires a thorough understanding of the missing data mechanism.

6.2.3 Processing of missing data

Many statistical analyses are predicated on the assumption of complete data for all variables included in analysis. As such, it is essential to minimise the impact of missing data within datasets. There are two divergent approaches to minimise the impact of missing data. The first approach is to omit cases with missing data from calculations; the second is to impute values for missing data. These approaches are further expanded below.

6.2.4 Omission methods

There are two approaches to omission of missing data - *list wise deletion* (also called complete case analysis), and *pair wise deletion* (also called available case analysis).

6.2.4.1 List wise deletion

In list wise deletion, all cases with missing data are removed from the dataset. The major advantage of list wise deletion is its simplicity. However, this approach can significantly reduce the size of the sample, typically 30-50%, on which statistical analyses are performed.[146] The smaller sample size in turn reduces the power of any analysis undertaken,[147] widens confidence intervals,[148] and can bias many parameter estimates.[149] Some argue that list wise deletion should only be considered when the proportion of missing data is below 5% and only if the missing data are MCAR.[146 147 150 151]

For example, **table 6.1** lists seven cases in a population, four of which have a missing data point. Under list wise deletion only the three cases with complete data would be available for analysis.

6.2.4.2 Pair wise deletion

In pair wise deletion, analysis is performed on all cases for which the variables of interest are present. Consequently, any given case may contribute to some analyses but not to others.[152] The major advantage of this approach is that it utilises as much of the available data as possible for analysis. The disadvantage of this approach is that each analysis is performed on a different sample, indeed sample sizes may vary considerably as a result.[153] A major concern with pairwise deletion is that there is no basis for estimating standard errors.[147 153] Pairwise deletion is not recommended as a general solution for missing data.[147]

For example, **table 6.1** lists seven cases in a population, four of which have a missing data point. Under pairwise deletion, three cases would contribute to all analyses while all cases would contribute to some analyse. Rows two and four cannot contribute to analyses requiring weight as an input, but can contribute to analyses requiring gender or age as the inputs.

Table 6.1 List wise and pair wise deletion

List wise deletion			Pair wise deletion		
Gender	**Age**	**Weight**	**Gender**	**Age**	**Weight**
M	27	96	M	27	96
~~M~~	~~27~~	~~NA~~	M	27	~~NA~~
~~F~~	~~NA~~	~~67~~	F	~~NA~~	67
~~M~~	~~45~~	~~NA~~	M	45	~~NA~~
F	33	58	F	33	58
~~NA~~	~~50~~	~~76~~	~~NA~~	50	76
F	22	80	F	22	80

6.2.5 Imputation methods

6.2.5.1 Last observation carried forward

Last observation carried forward methods are applicable to longitudinal studies only. This method imputes the missing values with the last recorded observation for the individual. Implicit with this method is the assumption that there has been little change in the individual since the last observation, which may well be an unrealistic assumption.[154] If the assumptions are unrealistic then the resulting estimations will be biased.

6.2.5.2 Single imputation

Single imputation is a simple method to replace missing values with estimated ones. Predictive mean matching (PMM) is the most common single imputation method. PMM entails substitution of the missing observation with the variable mean. It is both simple to implement and preserves the available data. However, it also skews the distribution of the variable toward the mean, and consequently biases estimates of both variances and covariances and impacts on correlations between variables.[148] The effect is worse the larger the proportion of missing data.[146] Single imputation methods lead to an underestimation of standard errors and consequently an overestimation of test statistics.[148] The primary reason for this is due to the imputed values being entirely determined by observed data without including any consideration for error in their estimation.[144] Single value imputation is not recommended as a general solution for missing data.

6.2.5.3 Deterministic regression imputation

Deterministic regression imputation employs a regression model to estimate values missing from an observation based upon variables that are present in the observation. Usually, linear regression is used for numeric variables, whereas for categorical data logistic regression may be used.[155] For example, using cases with complete data, a regression model is developed describing how weight can be estimated from age, height and waist measurement. For those cases with missing weight data, the same regression model is employed to predict weight based upon age, height and waist measurements within the observation. A disadvantage of regression model imputation is the failure to include an error term within the regression model, as such the imputed estimate will always be located on the regression line. This implies a perfect correlation between the predictors used and the variable with missing data - there is no residual variance.

Deterministic regression imputation thus overestimates correlations between predictors and outcomes, and underestimates the variances and covariances.

6.2.5.4 Stochastic regression imputation

Stochastic regression imputation extends the deterministic regression method by including a residual error term, to reduce the bias inherent in deterministic regression. This residual error term is normally distributed with a mean of zero and a variance equal to the residual variance from the regression of the predictor on the outcome.[152] This way the variability in the data is preserved and parameter estimates are unbiased provided the data are MAR.[152] However, the standard error tends to be underestimated, because the uncertainty about the imputed values is not included, which increases the risk of type I errors.[150]

6.2.5.5 Hot-deck and cold-deck imputation

Hot-deck imputation pairs observations containing missing values, with similar observations containing the required data, and then substitutes the value recorded from the observed case into observation with the missing value.[156] There are two general approaches to identifying 'similar' cases that will be used to provide substitute data, the nearest neighbour approach and the pattern matching approach.

The nearest neighbour approach (also called a distance function approach), imputes the missing value with the score of the case with the smallest squared distance statistic to the case with the missing value. The pattern matching approach stratifies the dataset into separate homogenous groups. The missing data value is then randomly drawn from the cases of the equivalent homogenous group.[157] Hot-deck imputation thus replaces the missing data with plausible scores that preserve the variable distribution, however the approach underestimates both standard errors and variability.[156]

Cold-deck imputation is far less common than hot deck imputation, but is essentially the same as hot-deck implementation with the only difference being the source of comparison data.[158] In hot-deck imputation the missing values are imputed from the same dataset, whereas in cold-deck imputation the missing values are imputed from a different dataset.[159]

6.2.5.6 Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE) is one of the most commonly employed imputation techniques. A model is defined that maximises the probability of observing the data that is present. In other words, MLE generates the model that is most consistent with the distribution of data that is observed (the data that is not missing). Missing values are neither replaced nor imputed, rather the missing data is handled within the analysis model. Maximum Likelihood Estimation is commonly observed in clinical practice, for example, a patient with a history of asthma is experiencing breathing difficulties, is struggling to talk and appears to be tiring. Although there are several possible causes of breathing difficulty (asthma, pulmonary embolism, pneumonia to name but a few), the observable data suggest the most likely cause is an acute asthma attack.

6.2.5.7 Multiple Imputation (MI)

A major flaw of deterministic regression imputation is that the imputed values always lie directly on the regression line, however real data seldom lies directly on the regression line. The deviation from the regression line represents the variance in the data.[144] Stochastic regression attempts to address this failing. Multiple imputation (MI) takes a further step and generates multiple datasets that include a term to account for variance of estimates. Consequently bias is reduced and estimates are more reliable.[160]

The key to MI lies in creating multiple versions of the dataset. MI seeks to replace missing values with plausible values, rather than estimates of the missing value,[127] and so generates multiple versions of the dataset, each of which will contain different imputed values that have been substituted for the missing values.[144]

The multiple imputation process contains three phases - imputation, analysis and pooling.[127 143 161]

Imputation phase

During imputation, multiple data sets are generated each of which contain different imputed values. The imputed values are drawn from the posterior probability distribution of the observed data[144] - thus multiple imputation is based on a bayesian approach.[162] The imputed values are estimated by applying a Gibbs sampling algorithm, a special *Markov Chain Monte Carlo (MCMC)*

technique, to the means and covariances of observed data; thereafter a residual term is added to the imputed value.[160] This process is iterated several times, updating the regression parameters following each iteration, resulting in different imputed values each time. After a defined number of iterations, the generated dataset is stored, and iterations re-commence. This process continues until the required number of imputed datasets is reached.

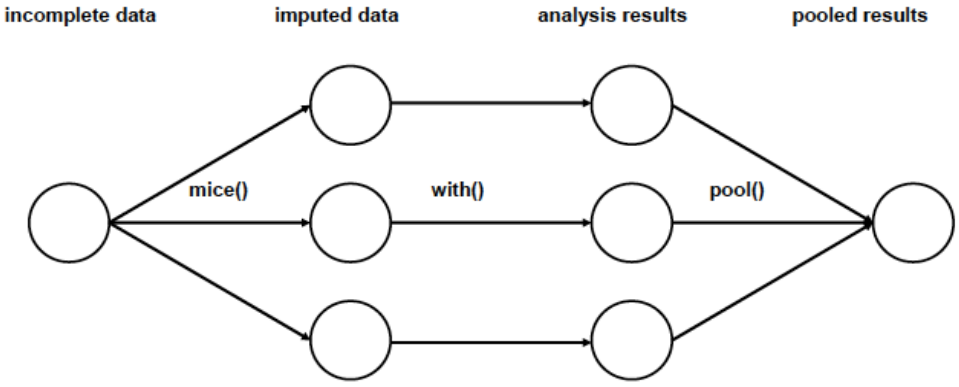
Analysis phase

During the analysis phase, standard statistical methods are applied to each of the imputed datasets.[163]

Pooling phase

During the pooling phase, statistical results generated during the analysis phase are combined into a single set of results. Statistical estimates are pooled using Rubin's rules,[164] which account for variability between the imputed datasets, reflecting the uncertainty associated with the missing values.[163] Parameter estimates are summarised by taking the average over the parameter estimates from all imputed datasets. The standard errors are pooled by combining the within imputation variance and the between imputation variance.[160]

Figure 6.1 Schematic of multiple imputation



Reproduced from Van Buuren *et al*[165]

Although MI fully accounts for uncertainty in predicting the missing values, we can never know the true values of the missing data.[163]

6.3 Method

Statistical analyses were performed using R Studio (Version 0.99.903). Missing data were processed by multiple imputation using the `MICE` package (Version 2.25). The `MICE` algorithm implements Multiple Imputation by Chained

Equations (MICE) with a Fully Conditional Specification (FCS) to impute missing values. A unique statistical model is thus derived for each variable. Standardised processing methods are provided for continuous data (predictive mean matching, normal), binary data (logistic regression), unordered categorical data (polytomous logistic regression) and ordered categorical data (proportional odds).[165]

Multiple imputation with `MICE` requires consideration of several factors to ensure appropriate specification of the imputation model.

6.3.1 Multiple imputation

Assessment of missing data - establish which data are missing, the types of variables with missing data and determine if there is any pattern to missingness.

6.3.1.1 Classification of missing data

Determine if the data are MCAR, MAR or MNAR. The likelihood that the data are MCAR is very low. MICE traditionally imputes data that are MAR. These methods are also valid for data that are MCAR. Although MICE can impute data that are MNAR, to successfully impute MNAR data requires additional assumptions that will influence the imputed data generated. Most often data are MAR.

6.3.1.2 Determine the imputation method for each imputed variable

In a Fully Conditional Specification (FCS) a unique model will be created to impute missing data for each variable i.e. there will be a model for each variable requiring imputation of missing data. The imputation method will be informed by variable type (see **table 6.2**).

6.3.1.3 Selection of variables to be included in the imputation model

Generally, the greater the number of variables used to impute missing data the more reliable the imputation is likely to be. However, some variables will contribute very little toward the model. Including many variables that contribute little to the model creates an unwieldily model, increases complexity and extends computation time with little benefit. Restricting imputation models to relevant variables optimises the imputation process.

Table 6.2 Imputation methods

Method	Description	Scale type	Default
pmm	Predictive mean matching	numeric	Yes
norm	Bayesian linear regression	numeric	
norm.nob	Non-Bayesian linear regression	numeric	
mean	Unconditional mean imputation	numeric	
2l.norm	Two level linear model	numeric	
logreg	Logistic regression	factor, 2 levels	Yes
polyreg	Polytomous (unordered) regression	factor, >2 levels	Yes
lda	Linear discriminant analysis	factor	
sample	Random sample from the observed data	any	

6.3.1.4 Decide whether to impute variables that are functions of other (incomplete) variables

Where variables are functions of included variables, for example the Glasgow Coma Score (GCS) is calculated from its three constituent parts (eye, verbal, motor response), it is possible to specify this relationship within the imputation model to ensure consistency, alternately one can omit the GCS variable and calculate it following imputation of the constituent parts.

6.3.1.5 Determine the visit sequence

The order in which variables are imputed (visit sequence) can have an impact on other variables, particularly in the case of variables that are functions of other variables. It may also impact convergence of the imputation algorithm.[165]

6.3.1.6 Determine the number of imputations

Alison[144] reports that early advocates of MI believed that three to five imputed data sets were sufficient. Schafer[161] later suggested that there was little to be gained by using more than five to ten imputed data sets, as the loss in efficiency was small even with 50% missing data. More recent debate has highlighted that a minor loss in efficiency does not necessarily equate to a minor loss in accuracy of standard error estimates, confidence intervals, and *p*-values.[144]

Graham *et al*[166] addressed the question of number of imputations required with respect to loss of power for hypothesis testing. They suggested that if a

researcher was willing to accept a loss of 1% in power, then 20 imputations were required for for 10% to 30% missing information, and 40 imputations would be required for 50% missing information. Bodner[167] and White *et al*[168] approach the question from a different perspective, they estimated the Monte Carlo error in reported *p*-values. Although they adopted different statistical approaches to the question, they arrived at similar conclusions, resulting in the current "rule of thumb" for number of imputations required, namely, *the number of imputations should be slightly higher than the percentage of cases that are incomplete*. Thus, if 37% of cases have missing data in one or more variables then 40 imputed data sets are required.

6.3.1.7 Choose the number of iterations

The number of iterations play an important role in assessment of convergence. As previously mentioned in *Section 6.2.5,* imputed values are estimated by applying a Gibbs sampling algorithm, a special *Markov Chain Monte Carlo (MCMC)* technique, to the means and covariance of observed data. This process is iterated several times, resulting in different imputed values for each iteration. The mean and variance of the imputations can be plotted to help assess for convergence of the Gibbs sampling algorithm. On a practical level, the greater then number of iterations the more data points can be plotted to visually assess convergence.

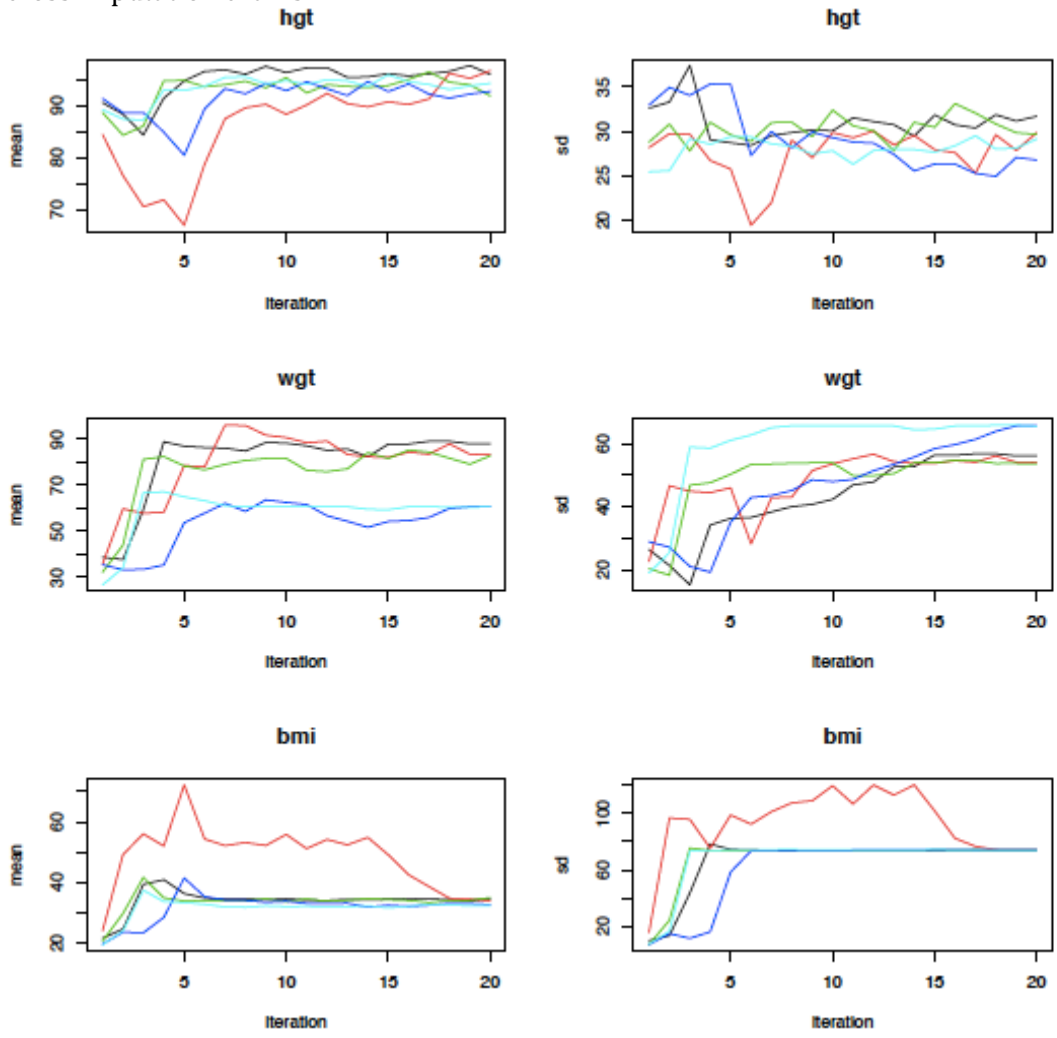
6.3.1.8 Assess imputed data

Successful MI with MICE is dependent upon convergence of the Gibbs sampling algorithm, however there is no clear-cut method for assessing whether the Gibbs sampling algorithm has converged. Technically, convergence is diagnosed when the variance between different sequences is no larger than the variance with each individual sequence.[127] Practically this can be assessed by plotting the mean and variance of the imputations against the iteration number. On convergence, the plots should be freely intermingled with each other, without showing any definite trends. **Figure 6.2** demonstrates unhealthy mixing of the imputation chains, where the plotted chains fail to mingle. **Figure 6.3** demonstrates healthy convergence with adequate mixing of the plotted chains.

Alternatively, the R-hat convergence statistics, calculated using the `MICEAdds` package (Version 1.9-0), can be calculated for each imputed variable. The R-hat convergence statistics compare the variance between chains to the variance

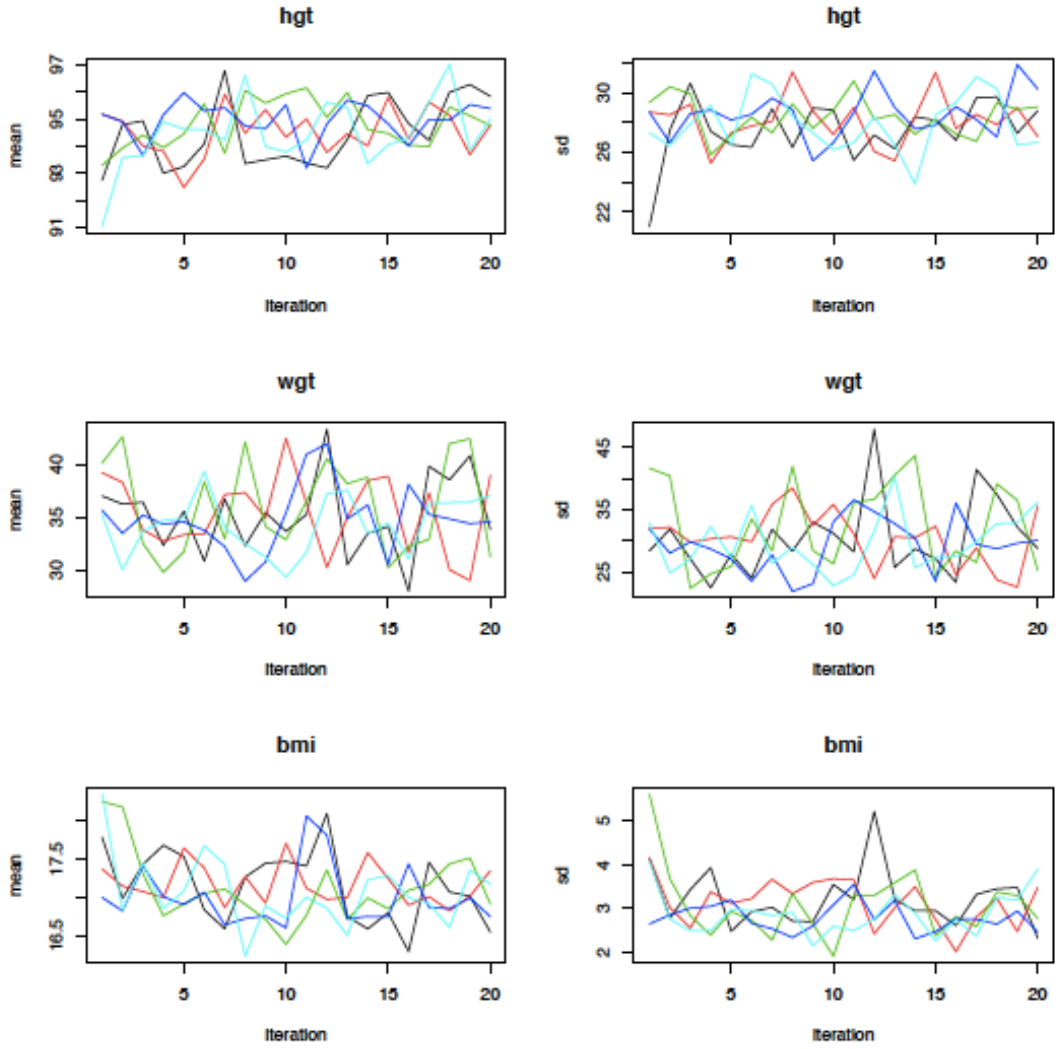
across chains. Values closer to 1.0 indicate little is to be gained by running the chains longer (more iterations not required), and in general, values greater than 1.1 indicate that the chains should be run longer (more iterations required).[169] Once convergence has been confirmed, inspection of imputed data is required. For large datasets, this is commonly achieved by inspection of density plots for each imputed variable to confirm the imputed data are consistent with the observed data.

Figure 6.2 Unhealthy convergence demonstrated by lack of mixing across imputation chains



Reproduced from van Buuren[127]

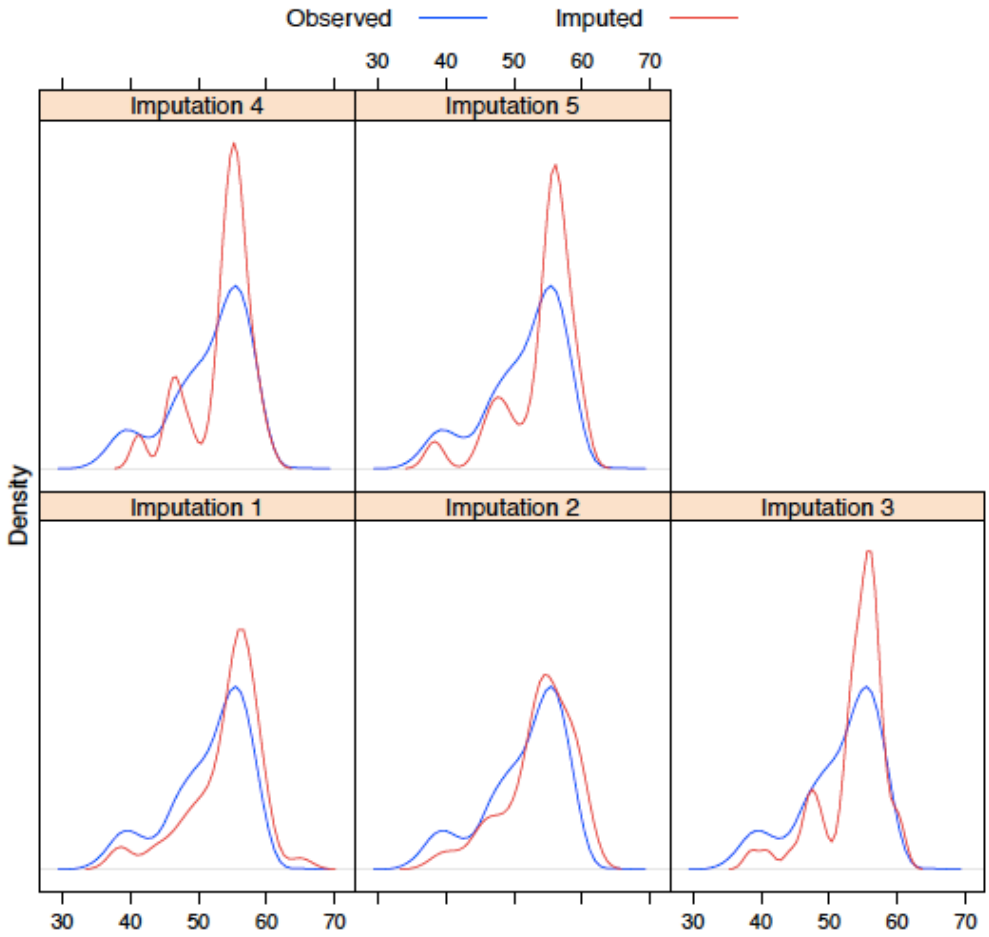
Figure 6.3 Healthy convergence demonstrated by mixing across imputation chains



Reproduced from van Buuren[127]

The MAR assumption can never be truly confirmed from the observed data - this is because we do not know what the true missing values are. However, we can confirm that the imputed values generated are plausible. For each variable we have imputed, we plot densities of both the observed and imputed values to visually determine if the imputed values are reasonable. Differences in the densities between the observed and imputed values suggest a problem that requires further investigation. **Figure 6.4** demonstrates a plot of healthy imputation where the densities of imputed data are similar to the observed data.

Figure 6.4 Healthy Density Plots



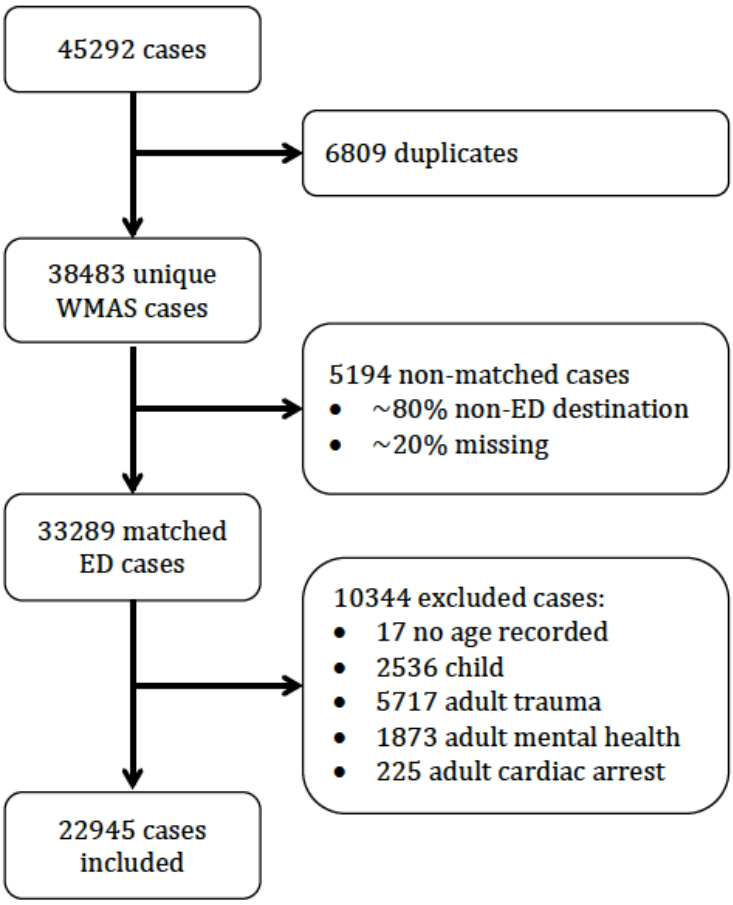
Reproduced from van Buuren[127]

6.4 Results

6.4.1 Removal of excluded cases

The previous chapter addressing record linkage confirmed that 33289 clinical records were successfully linked. However, these linked records include cases that were defined, *a priori,* as excluded case types. Excluded cases included age under 18 years, attendance for trauma, attendance for mental health issue and cases of cardiac arrest. Prior to imputing missing data, excluded cases were removed from the dataset. More reliable diagnostic data were available in the UHNS dataset, the discharge diagnoses stored in the variables `Diagnosis1`, `Diagnosis2`, `Diagnosis3`, `Diagnosis4` and `Diagnosis5` were utilised to removed excluded cases from the `WMAS` dataset. Ultimately 10344 cases were removed while 22945 cases were included for development of the prehospital sepsis screening tool (see **figure 6.1** for further details).

Figure 6.5 Included cases



6.4.2 Multiple imputation

6.4.2.1 Assessment of missing data

Examination of the `WMAS` data indicates they comprise 22945 observations each having 39 variables (see **table 6.3**).

Table 6.3 Composition of WMAS data

Variable	Type	Unique
WMASID	character	22945
UHNSID	character	22945
DispCat	factor	8
DispComp1	factor	30
DispComp2	factor	29
Imp	factor	6
Location	factor	3
Age	integer	88
Gender	factor	2
Ethnicity	factor	7
Resps	integer	53
SpO2	integer	58
Pulse	integer	201
SBP	integer	200
DBP	integer	141
Temp	numeric	84
BM	numeric	302
GCS_eye	ordered-factor	4
GCS_verbal	ordered-factor	5
GCS_motor	ordered-factor	6
GCS_sum	integer	13
Skin	factor	7
CBRT	ordered-factor	2
AVPU	ordered-factor	4
RPupilReact	ordered-factor	3
RPupilSize	integer	10
LPupilReact	ordered-factor	3
LPupilSize	integer	10
Oxygen	ordered-factor	2
Fluids	ordered-factor	2
UHNS_Infection	logical	2
UHNS_ED_Dr	ordered-factor	5
UHNS_NICE_Sepsis	ordered-factor	4
UHNS_OF_Sepsis	logical	2
UHNS_SOFA_Sepsis	ordered-factor	3
UHNS_SIRS_Sepsis	ordered-factor	3
UHNS_ED_Outcome	factor	89
UHNS_Died_In_ED	logical	2
UHNS_Died_In_Hospital	logical	2

Proportions of missing data are reported in **table 6.4**.

Table 6.4 Missing data WMAS

Variable	Missing (n)	Missing (%)
WMASID	0	0
UHNSID	0	0
DispCat	0	0
DispComp1	46	0.2
DispComp2	18869	82.24
Imp	224	0.98
Location	0	0
Age	0	0
Gender	0	0
Ethnicity	11051	48.16
Resps	59	0.26
SpO2	163	0.71
Pulse	130	0.57
SBP	393	1.71
DBP	427	1.86
Temp	3540	15.43
BM	5238	22.83
GCS_eye	302	1.32
GCS_verbal	302	1.32
GCS_motor	302	1.32
GCS_sum	302	1.32
Skin	2638	11.5
CBRT	3052	13.3
AVPU	49	0.21
RPupilReact	2177	9.49
RPupilSize	1575	6.86
LPupilReact	2104	9.17
LPupilSize	1490	6.49
Oxygen	0	0
Fluids	0	0
UHNS_Infection	0	0
UHNS_ED_Dr	0	0
UHNS_NICE_Sepsis	0	0
UHNS_OF_Sepsis	0	0
UHNS_SOFA_Sepsis	0	0
UHNS_SIRS_Sepsis	0	0
UHNS_ED_Outcome	0	0
UHNS_Died_In_ED	0	0
UHNS_Died_In_Hospital	0	0

Both DisComp2 and Ethnicity have very high levels of missingness. The variables BM, Temp, CBRT and Skin have moderate missingness. The remaining variables have fewer than 10% missing data (see **table 6.4**). van Buuren[127 165] suggests that variables with high missingness should be removed and should not be imputed, consequently DisComp2 and Ethnicity were removed from the WMAS dataset.

When assessing for any apparent patterns in the missing data we note that missingness for Glasgow Coma Score elements (GCS_eye, GCS_verbal, GCS_motor and GCS_sum) is identical, in addition there are similar proportions of missingness between systolic and diastolic blood pressure (SBP and DBP), pupil reactions (RPupilReact and LPupilReact) and pupil size (RPupilSize and LPupilSize). It can be useful to look for patterns in the missing data by examining a graphical output of missingness.

Figure 6.6 is a bar chart of missing data. Variables are ordered left to right in terms of proportion of missing data. Seven variables have greater than 5% missing data, while four have greater than 10%. Most variables have fewer than 2% missing. **Figure 6.7** shows missingness for each case with at least one missing value (separation between observations are too small for this to be clearly visible), blue is observed data while red represents missing data. Four patterns can be observed to suggest there may be a relationship between the missing data. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) appear to be missing together, although this is not always the case. Elements of the Glasgow Coma Score (GCS_eye, GCS_verbal, GCS_motor) appear to be missing together. It also appears that the variables associated with pupil reactions (LPupilReact and RPupilReact) and pupil size (LPupilSize and RPupilSize) appear to be missing together.

Figure 6.6 Bar chart of missing data

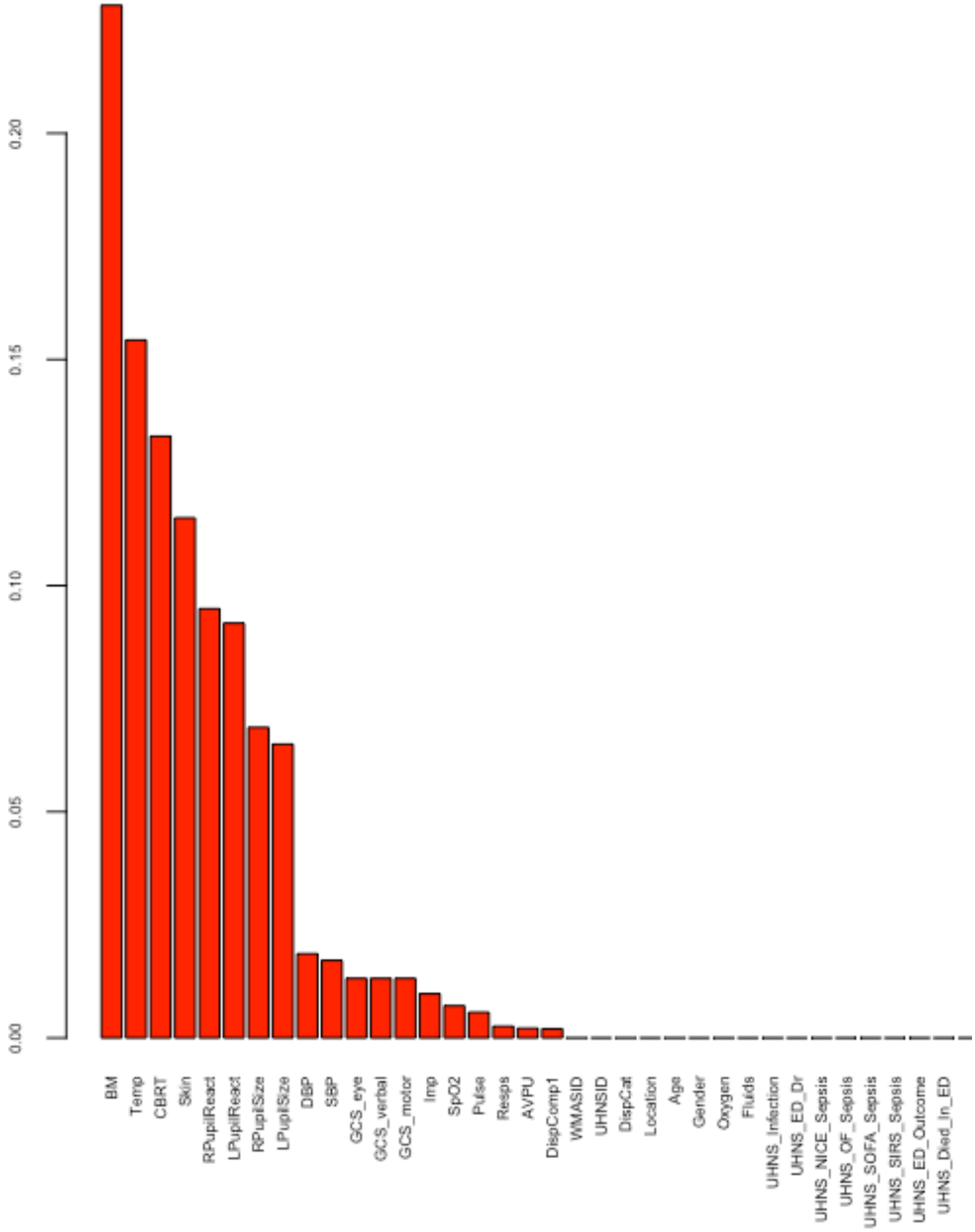
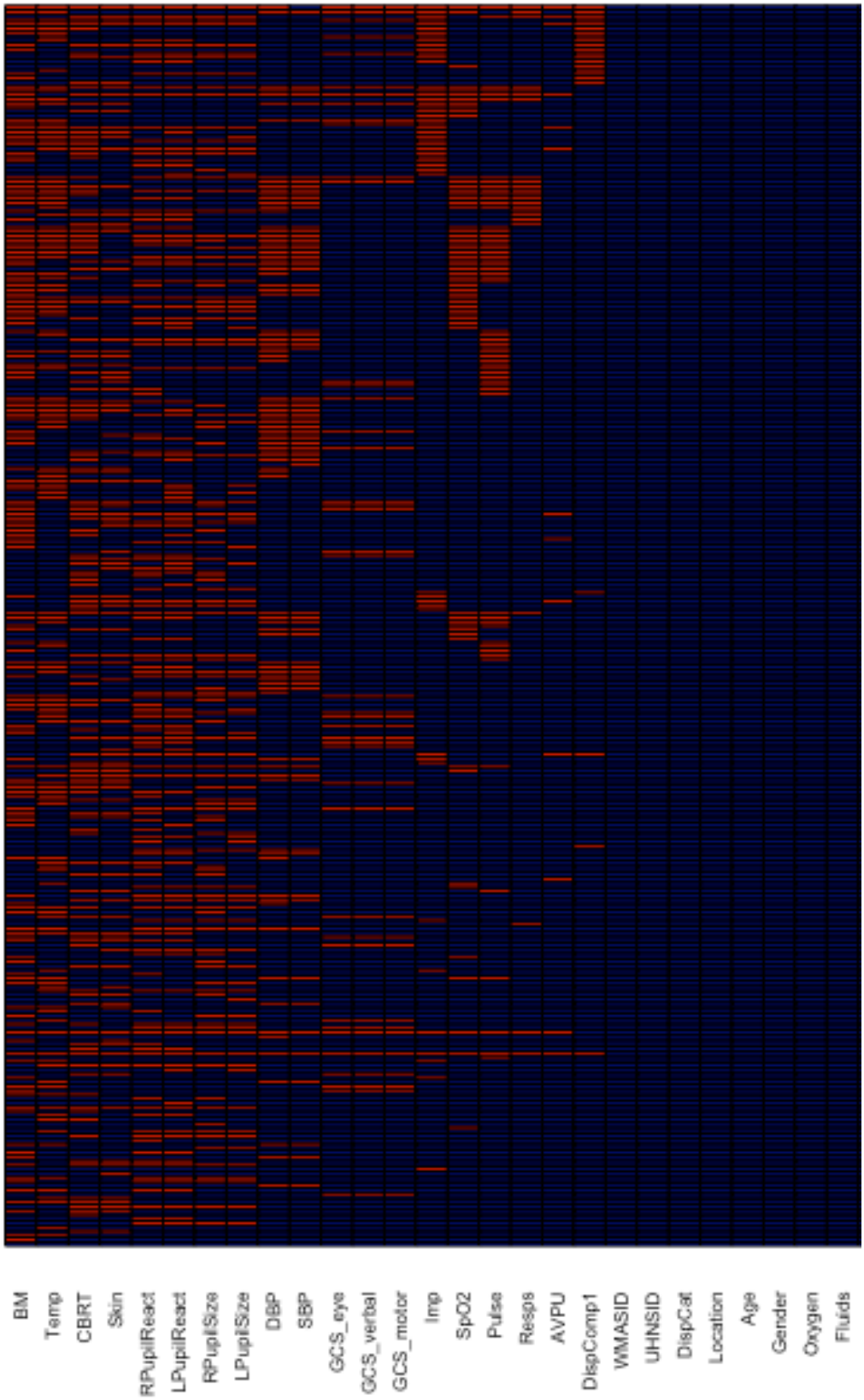


Figure 6.7 Missingness of data



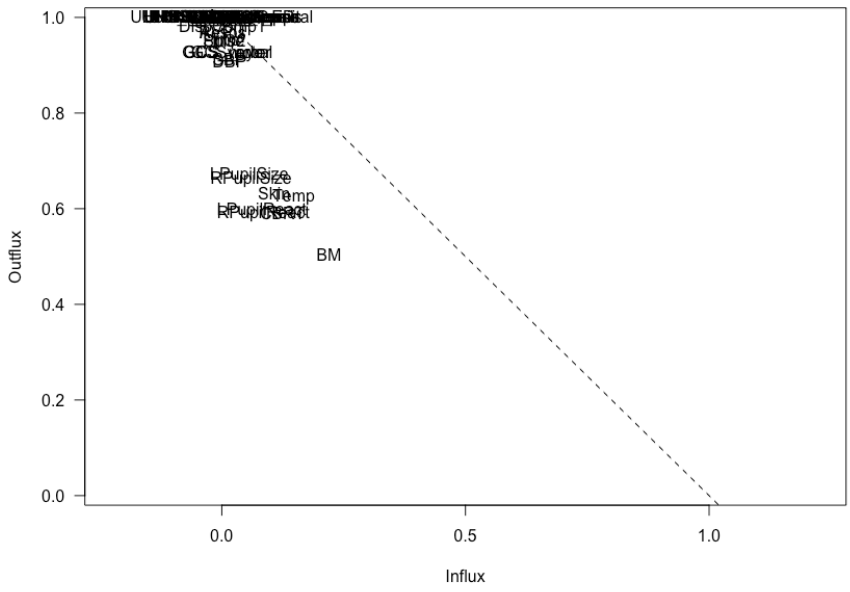
6.4.2.2 Classification of missing data

Although there appear to be relationships between some of the missing variables, when considered from the perspective of the clinician performing an assessment it becomes less likely that the data are MNAR. For example, a clinician fails to take a blood pressure, resulting in both `SBP` and `DBP` being missing - the two variables relate to a single assessment. Similarly, the variables `GCS_eye`, `GCS_verbal` and `GCS_motor` are elements of a single Glasgow Coma Score assessment, left and right pupil sizes (`LPupilSize` and `RPupilSize`) as well as reactions (`LPupilReact` and `RPupilReact`) are also a single assessment. Consequently, when considered as assessments performed, rather than individual variables, the missingness pattern disappears, and can no longer be considered MNAR. As a result, the data are either MAR or MCAR. We have no means of proving data are MCAR. As such the data can be assumed to be MAR.

6.4.2.3 Selection of variables to be included in the imputation model

We can assess which variables are amenable to imputation by generating a flux plot (see **figure 6.8**). Variables appearing in top left indicate little missing and very amenable to imputation, bottom right suggest high missingness and may generate unreliable imputations. Most variables are clustered in the top left indicating imputation should be successful. No variables appear in the bottom right suggesting no unreliable imputations will be generated. A number of variables are clustered slightly above mid-graph. These data will be imputed and success of the imputation closely examined.

Figure 6.8 Fluxplot of variables



Only variables that could conceivably be included in a screening tool are imputed. Several variables have no clinical value and will not be imputed (see **table 6.5**).

Table 6.5 Imputation of variables

Variable	Imputed/not imputed
WMASID	not imputed
UHNSID	not imputed
DispCat	not imputed
DispComp1	not imputed
Imp	imputed
Location	not imputed – no missing data
Age	not imputed - no missing data
Gender	not imputed - no missing data
Resps	Imputed
SpO2	Imputed
Pulse	Imputed
SBP	Imputed
DBP	Imputed
Temp	Imputed
BM	Imputed
GCS_eye	Imputed
GCS_verbal	Imputed
GCS_motor	Imputed
GCS_sum	Imputed
Skin	Imputed
CBRT	Imputed
AVPU	Imputed
RPupilReact	Imputed
RPupilSize	Imputed
LPupilReact	Imputed
LPupilSize	Imputed
Oxygen	not imputed - no missing data
Fluids	not imputed - no missing data

6.4.2.4 Decide whether to impute variables that are functions of other (incomplete) variables

The only variable that is a function of other variables is GCS_sum which is calculated from the sum of GCS_eye, GCS_verbal and GCS_motor scores. GCS_sum is initially removed as a variable, howevere it will be calculated from imputed values for GCS_eye, GCS_verbal and GCS_motor scores, and reinserted following completion of the multiple imputation.

6.4.2.5 Determine the imputation method for each imputed variable

`Mice` can analyse the variables present in a data frame and generate a matrix indicating which variables should be used to model missing data for each variable (see **table 6.6**). This matrix is modifiable by the user if they feel an important relationship has been omitted. For example, `Resps` is imputed utilising a model that includes `Imp`, `SpO2`, `Pulse`, `SBP`, `Temp`, `BM` and `CBRT`.

Table 6.6 Predictor matrix

	Imp	Resps	SpO2	Pulse	SBP	DBP	Temp	BM	GCS_eye	GCS_verbal	GCS_motor	Skin	CBRT	AVPU	RPupilReac +	RPupilSize	LPupilReac +	LPupilSize
Imp	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Resps	1	0	1	1	1	0	1	1	0	0	0	0	1	0	0	0	0	0
SpO2	1	1	0	1	0	0	1	1	1	1	1	1	1	1	0	0	0	0
Pulse	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
SBP	0	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0
DBP	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
Temp	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
BM	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GCS_eye	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0
GCS_verbal	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0
GCS_motor	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0
Skin	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
CBRT	0	1	1	0	0	0	0	0	1	1	1	0	0	1	1	0	1	0
AVPU	0	0	1	0	0	0	0	0	1	1	1	0	1	0	1	0	1	0
RPupilReact	0	0	0	0	0	0	0	0	1	1	1	0	1	1	0	0	1	0
RPupilSize	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
LPupilReact	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0
LPupilSize	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

The default methods for imputation were employed. See **table 6.7**.

Table 6.7 Imputation method

Variable	Imputation method
Imp	polytomous regession
Resps	predictive mean matching
SpO2	predictive mean matching
Pulse	predictive mean matching
SBP	predictive mean matching
DBP	predictive mean matching
Temp	predictive mean matching
BM	predictive mean matching
GCS_eye	polytomous regession
GCS_verbal	polytomous regession
GCS_motor	polytomous regession
Skin	polytomous regession
CBRT	logistic regession
AVPU	polytomous regession
RPupilReact	logistic regession
RPupilSize	predictive mean matching
LPupilReact	logistic regession
LPupilSize	predictive mean matching

6.4.2.6 Determine the visit sequence

There were no dependencies within the variables. As such, the default visit sequence (left to right) was used.

6.4.2.7 Determine the number of imputations

The number of imputations is dependent upon proportion of cases with missing data. There are 22945 cases of which 12517 are complete (54.6%), leaving slightly over 45% of cases with at least 1 missing data point. Consequently, based upon the rule of thumb requiring one imputation per 1% missing data, the number of imputations required was set at 50.

6.4.2.8 Choose the number of iterations

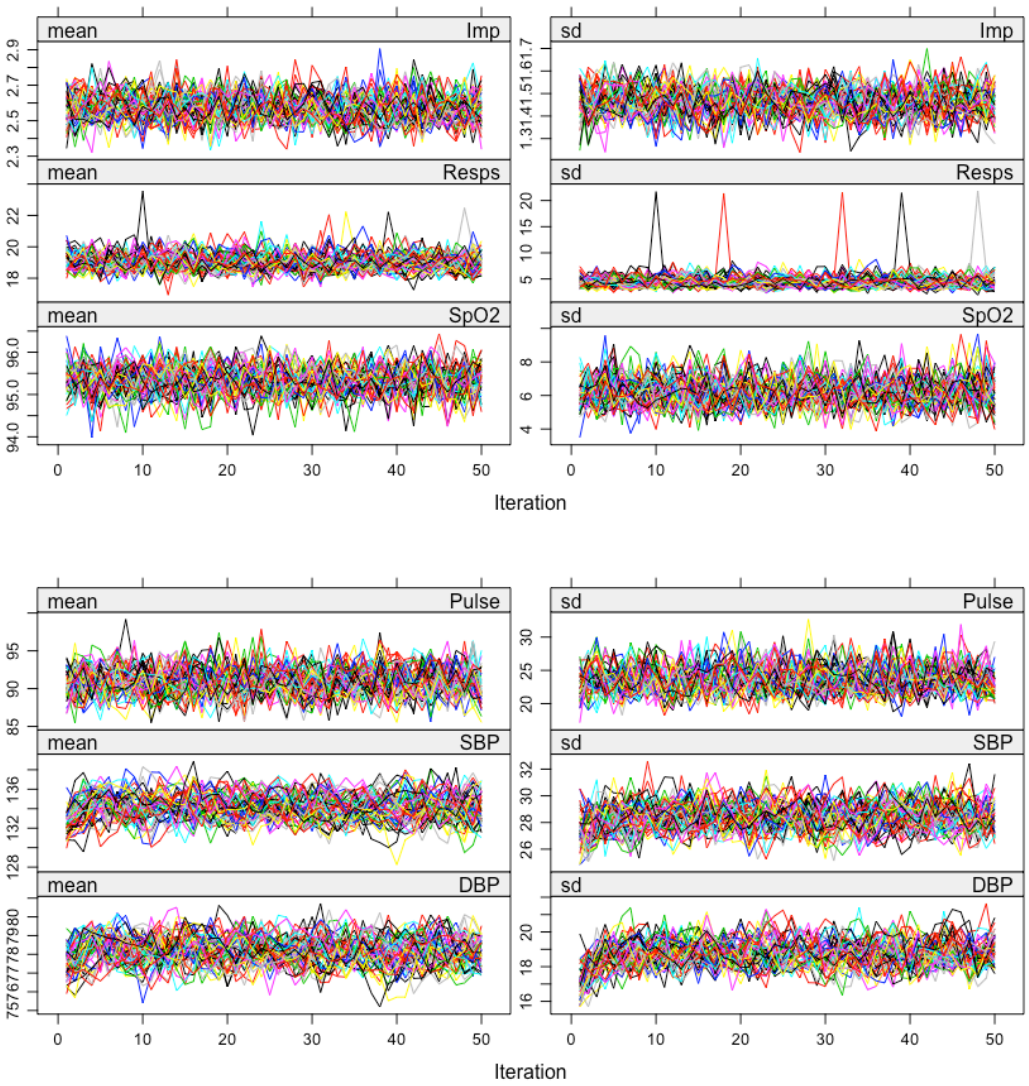
There are no reliable methods to prospectively determine the precise number of iterations that will be required to achieve healthy convergence, it is generally confirmed in retrospect. The default number of iterations in `mice` is five iterations. Plotting the five imputed data chains failed to demonstrate that convergence had been achieved, consequently the chain length was increased to

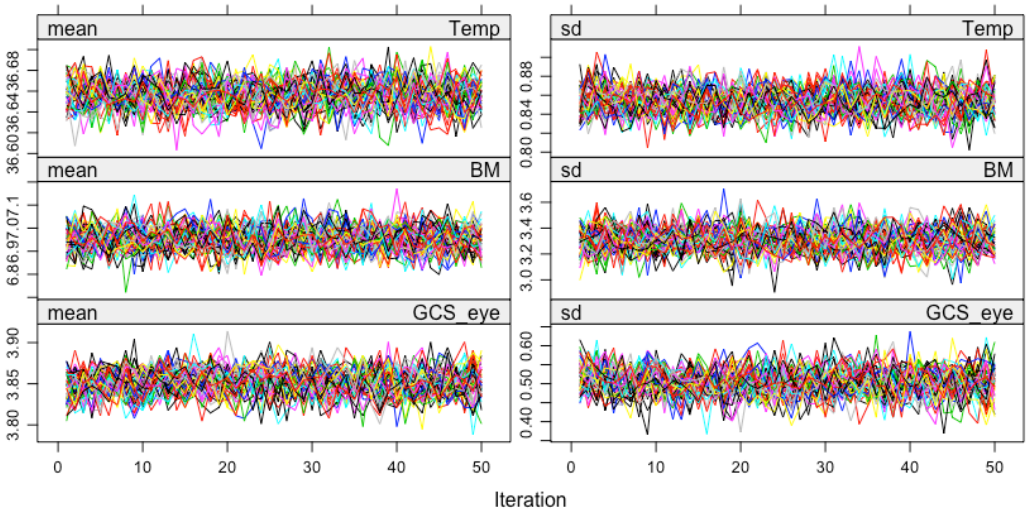
ten iterations, and the imputation plots were reassessed. This plot again failed to show convergence had been achieved. This process was repeated with 20 iterations, 30 iterations, 40 iterations and finally 50 iterations.

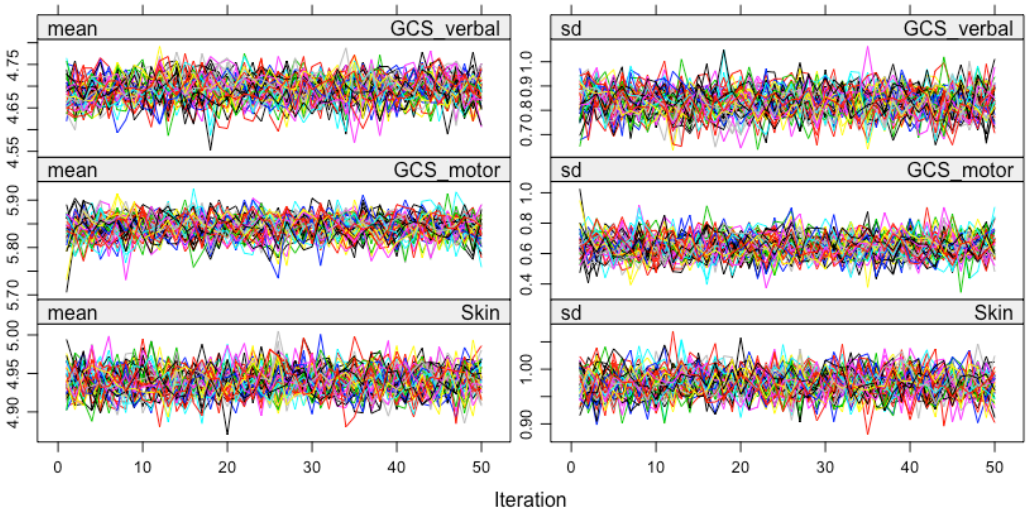
6.4.2.9 Assess imputed data

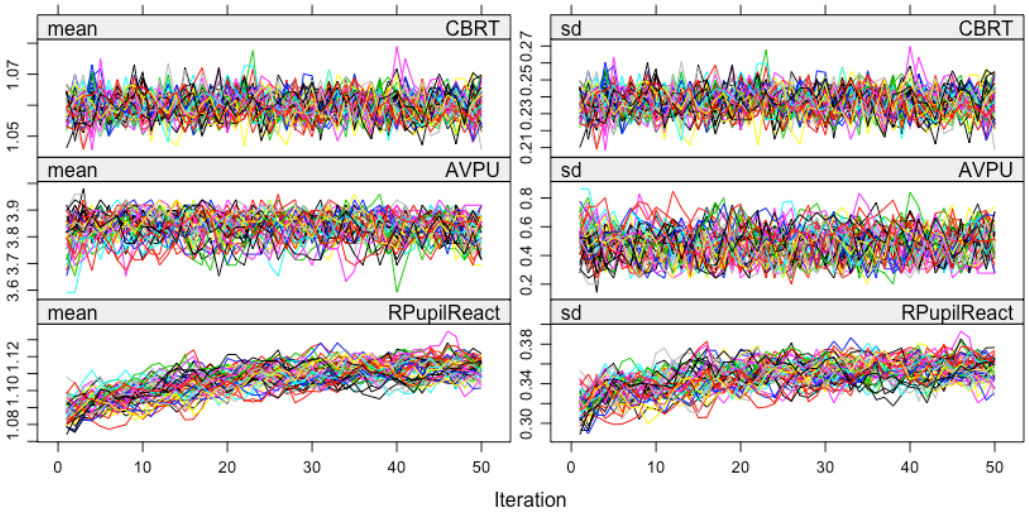
Convergence plots demonstrated excellent convergence, seen by healthy mixing of the chains, for all imputed variables except for `LPupilSize` and `RPupilSize`. See **figure 6.9**.

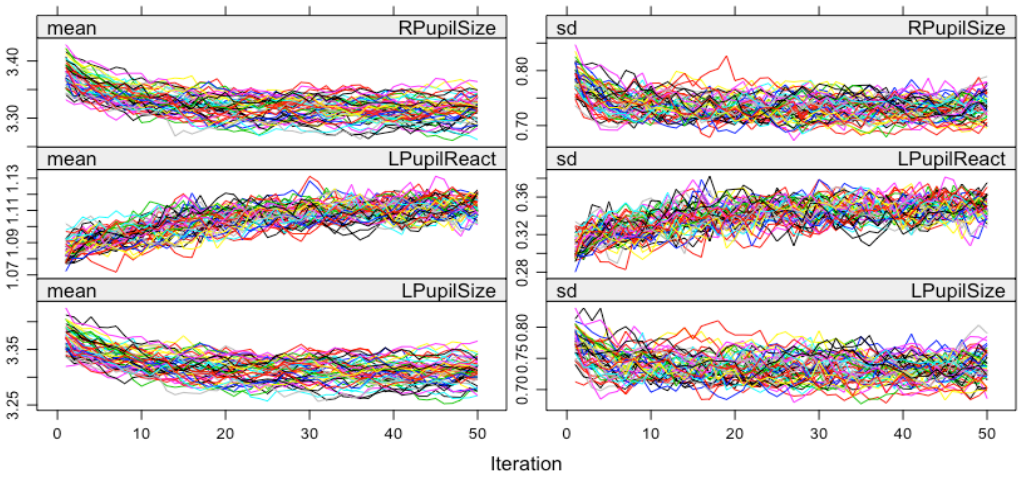
Figure 6.9 Convergence plots




mean
Temp
sd
Temp
mean
BM
sd
BM
mean
GCS_eye
sd
GCS_eye
Iteration


mean
GCS_verbal
sd
GCS_verbal
mean
GCS_motor
sd
GCS_motor
mean
Skin
sd
Skin
Iteration


mean
CBRT
sd
CBRT
mean
AVPU
sd
AVPU
mean
RPupilReact
sd
RPupilReact
Iteration



Rhat statistics indicate successful convergence of imputations for all imputed variables except `LPupilSize` and `RPupilSize`. See **table 6.8**.

Table 6.8 Rhat scores

Variable	% Missing	Mean	Variance
Imp	0.9762475	1.0016182	1.0047219
Resps	0.2571366	0.9991814	1.0523770
SpO2	0.7103944	1.0026238	1.0008604
Pulse	0.5665722	1.0037685	0.9977388
SBP	1.7127915	1.0248377	1.0073032
DBP	1.8609719	1.0253729	1.0076894
Temp	15.4281979	1.0011967	0.9988709
BM	22.8285029	1.0017126	1.0002574
GCS_eye	1.3161909	1.0092198	1.0081991
GCS_verbal	1.3161909	1.0126162	1.0070718
GCS_motor	1.3161909	1.0095771	1.0100301
Skin	11.4970582	0.9984954	0.9984608
CBRT	13.3013728	0.9993559	0.9993735
AVPU	0.2135542	1.0334403	1.0201990
RPupilReact	9.4879059	1.0446562	1.0445606
RPupilSize	6.8642406	1.2355947	1.2227659
LPupilReact	9.1697538	1.0489290	1.0491825
LPupilSize	6.4937895	1.2547893	1.2512539

Density plots for continuous variables indicate appropriate distributions across the imputed datasets were achieved. See **figures 6.10** to **6.18**.

Figure 6.10 Density plot `Resps`

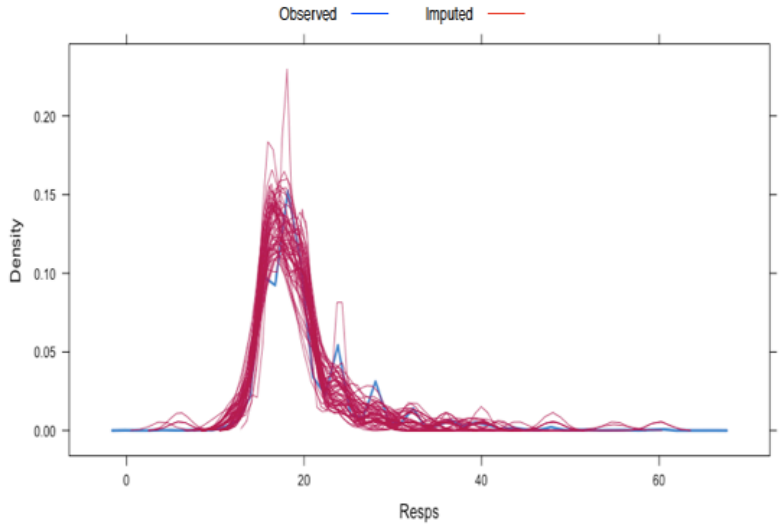


Figure 6.11 Density plot SpO_2

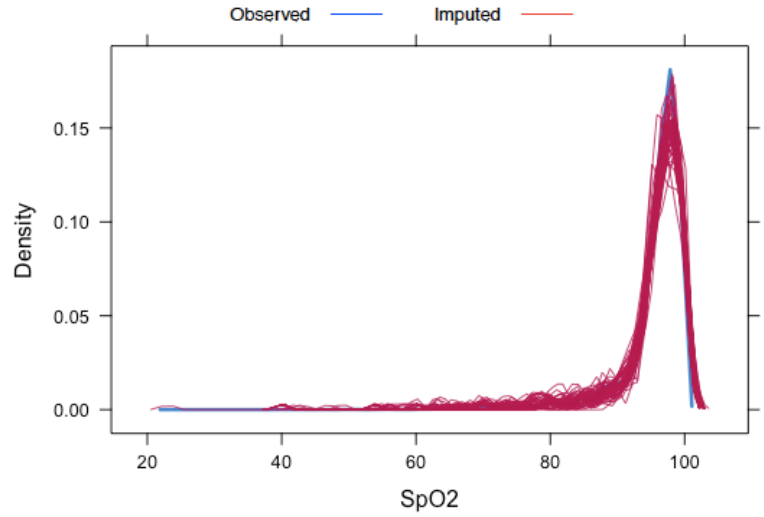


Figure 6.12 Density plot Pulse

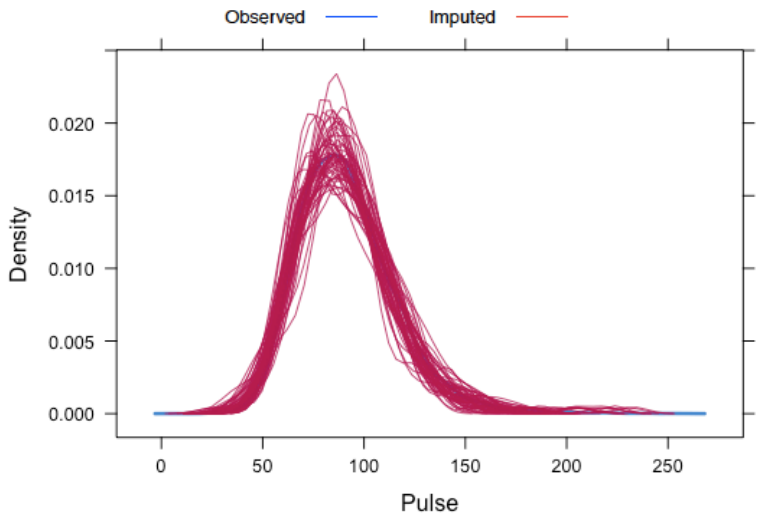


Figure 6.13 Density plot SBP

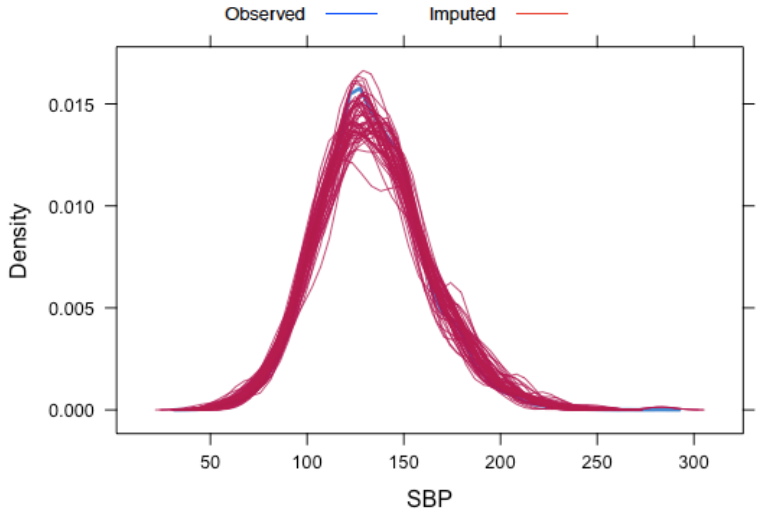


Figure 6.14 Density plot DBP

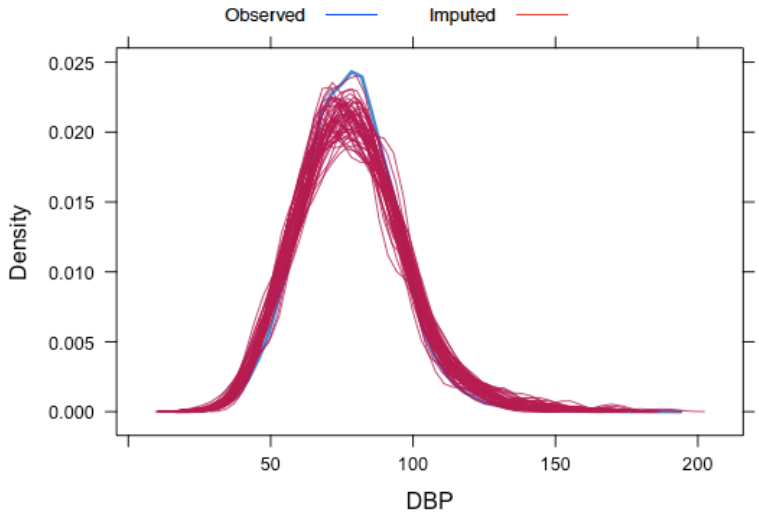


Figure 6.15 Density plot `BM`

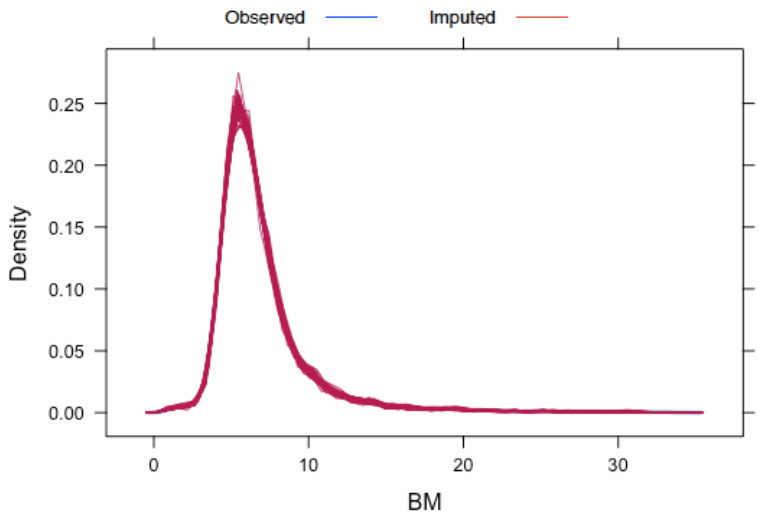


Figure 6.16 Density plot `Temp`

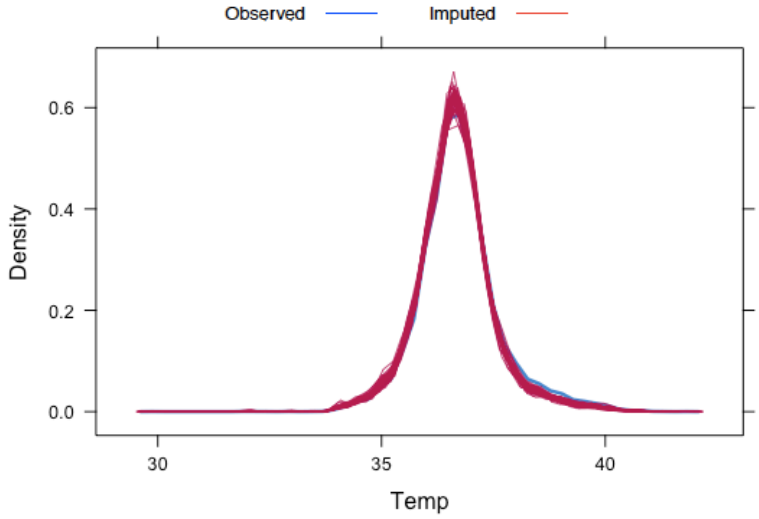


Figure 6.17 Density plot `LPupilSize`

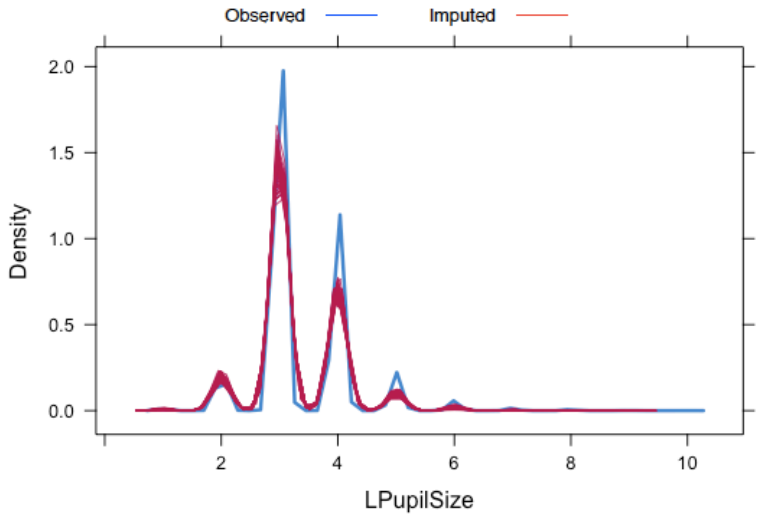
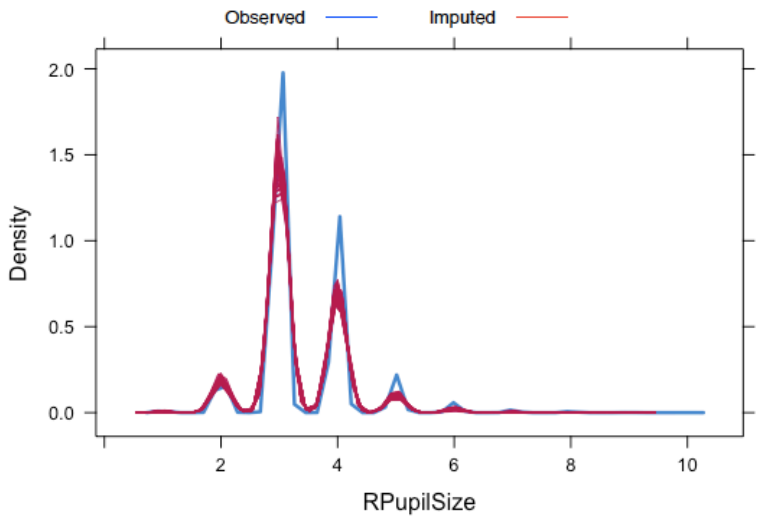


Figure 6.18 Density plot `RPupilSize`



Box and whisker plots for continuous variables also indicate appropriate distributions across the imputed datasets were achieved. See **figures 6.19** to **6.27**.

Figure 6.19 Box and Whisker plot `Resps`

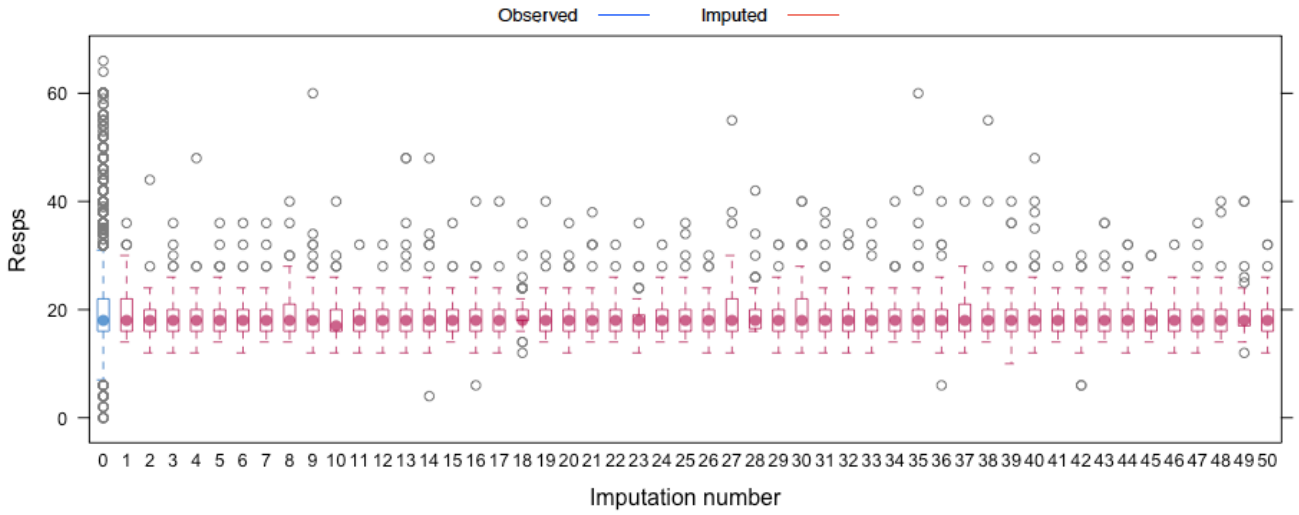


Figure 6.20 Box and Whisker plot SpO2

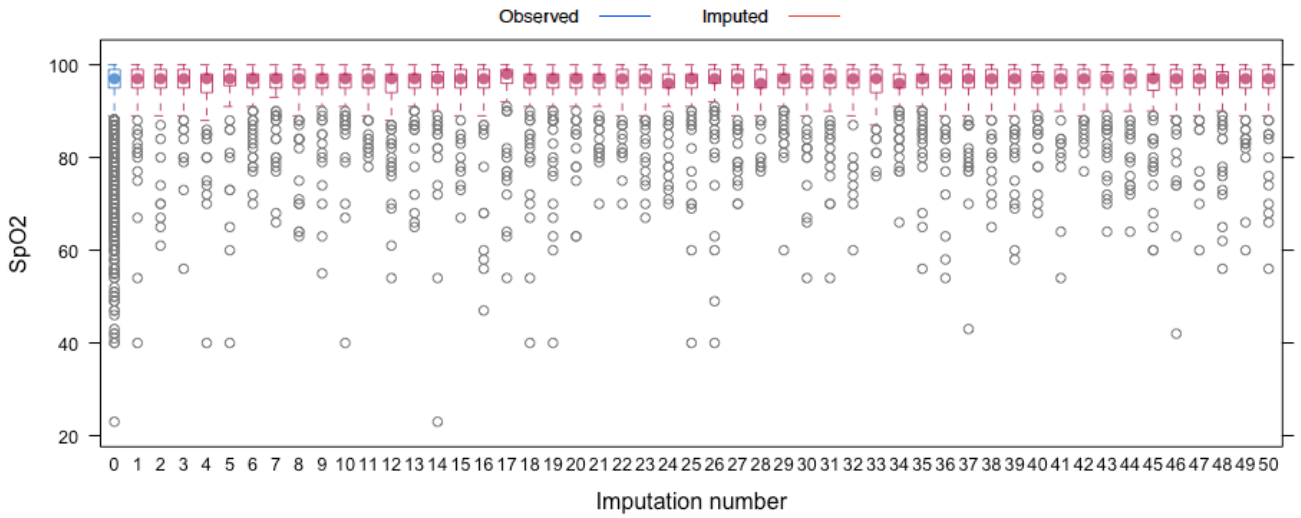


Figure 6.21 Box and Whisker plot `Pulse`

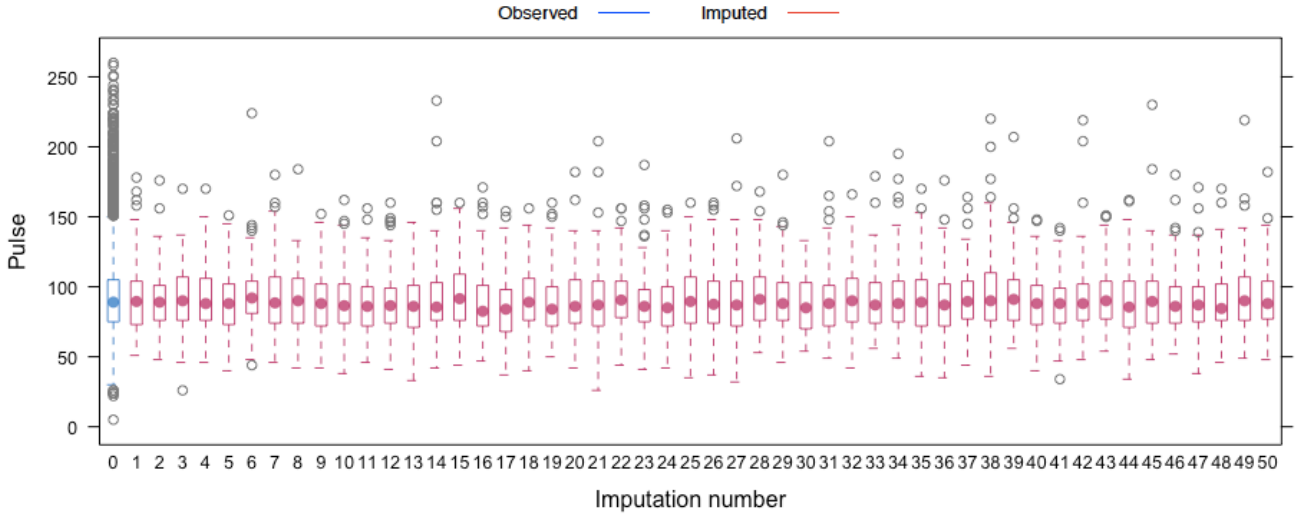


Figure 6.22 Box and Whisker plot SBP

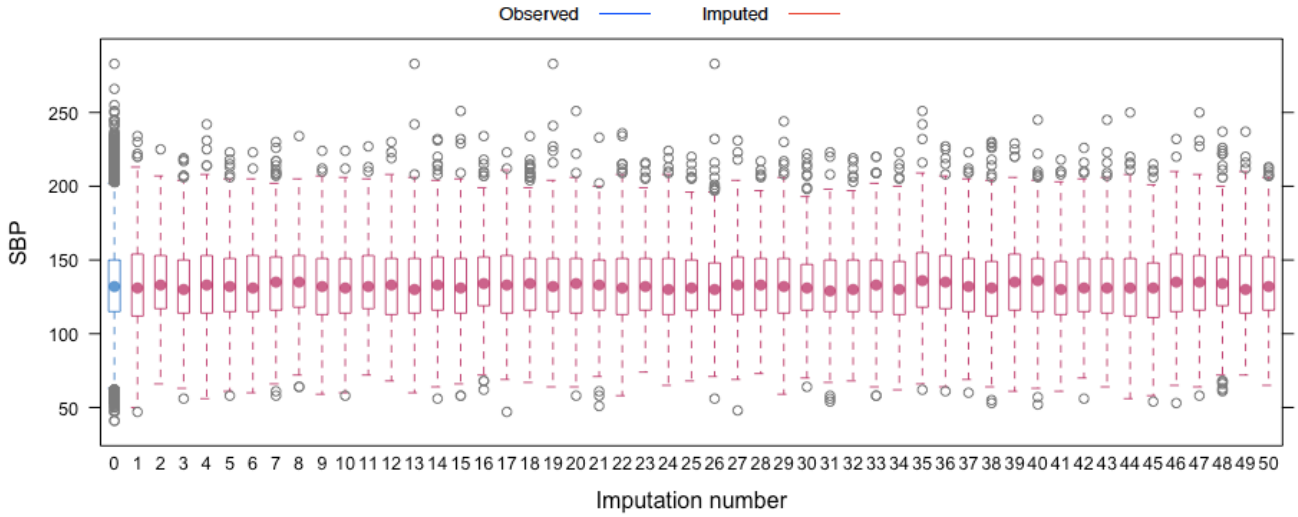


Figure 6.23 Box and Whisker plot DBP

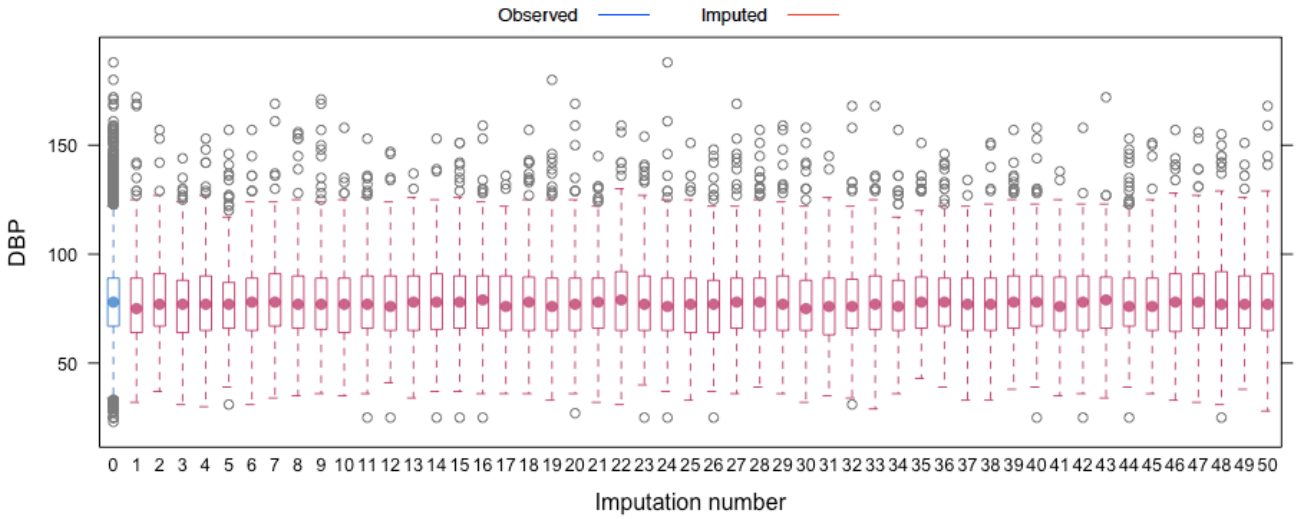


Figure 6.24 Box and Whisker plot BM

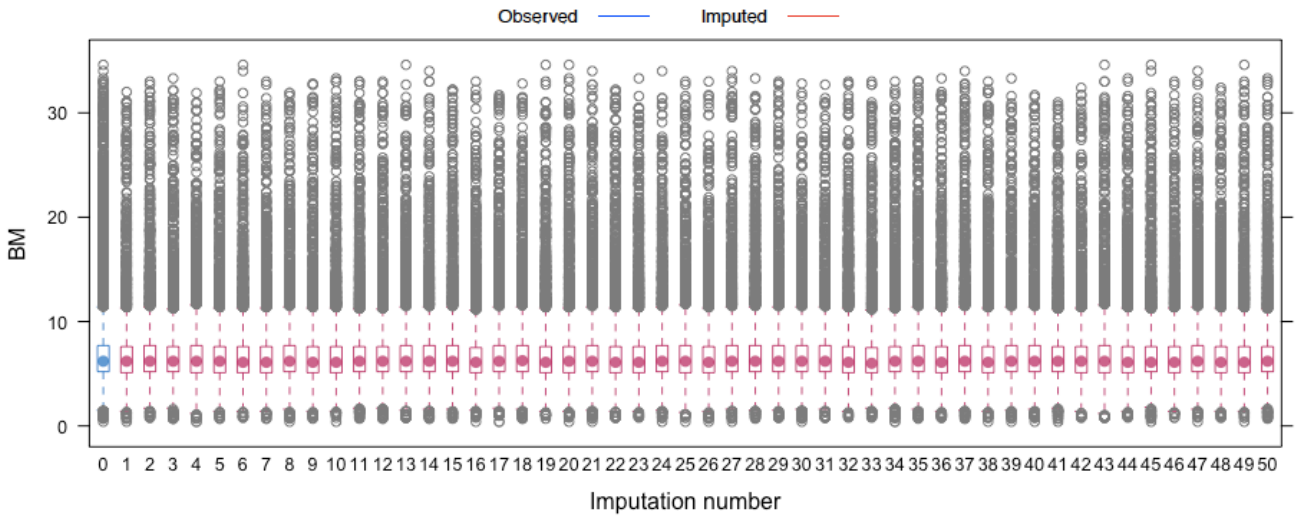


Figure 6.25 Box and Whisker plot `Temp`

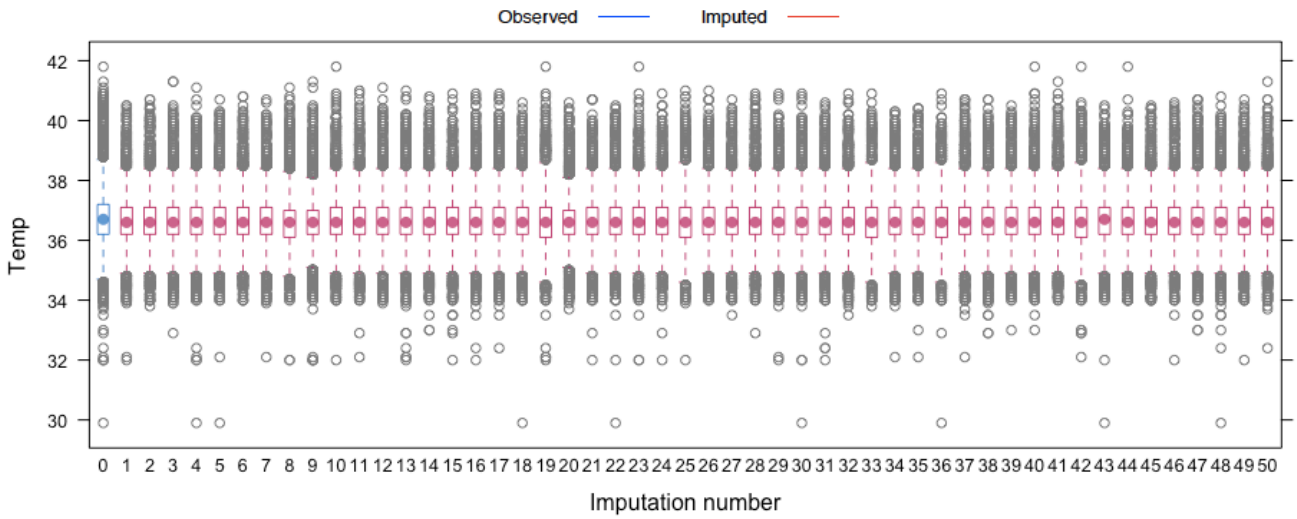


Figure 6.26 Box and Whisker plot `LPupilSize`

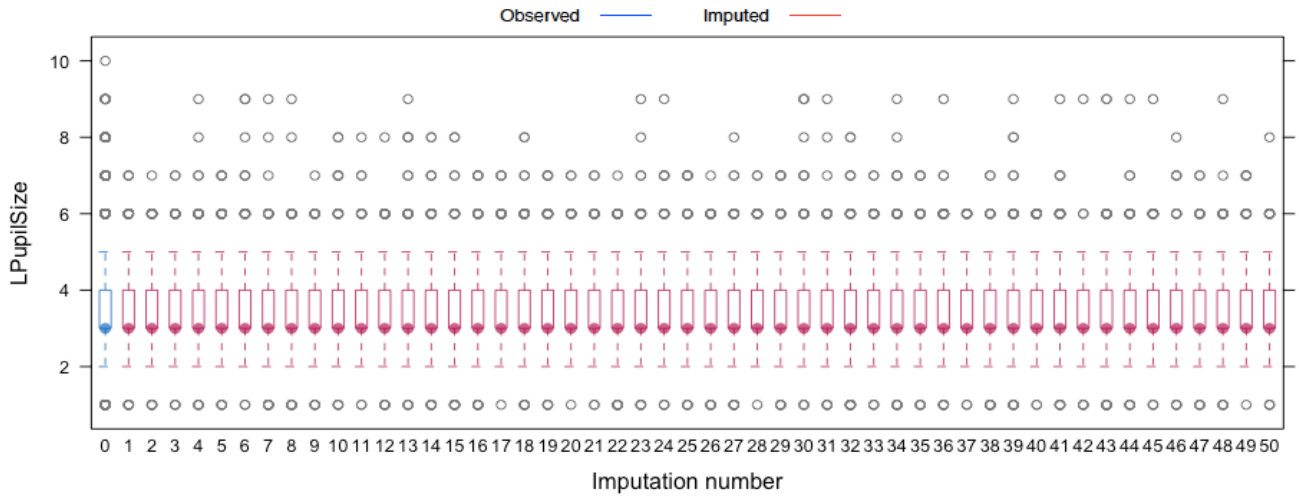
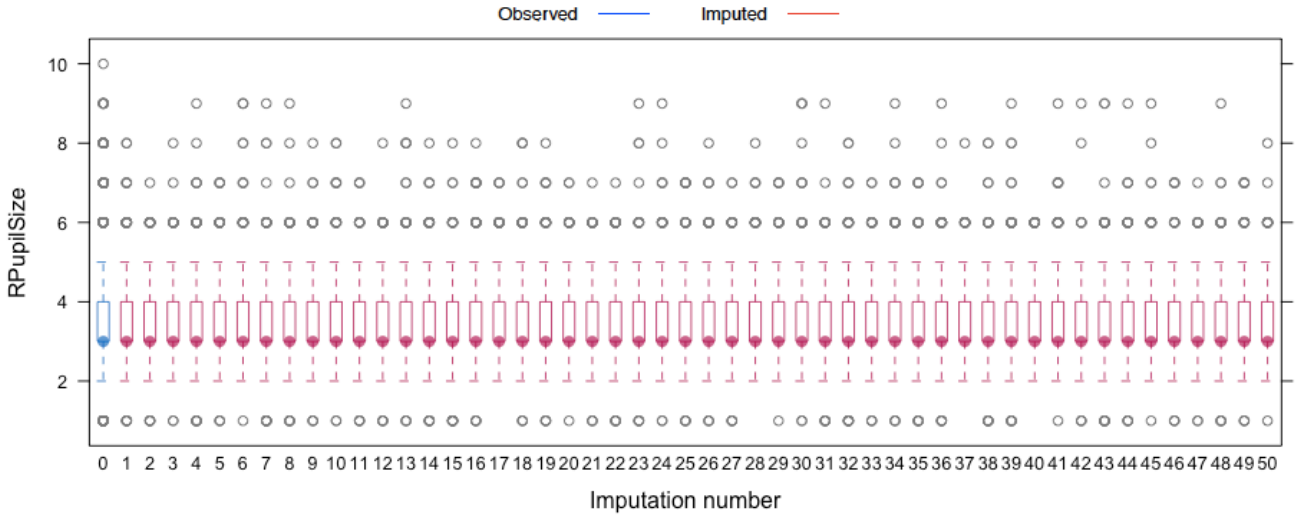


Figure 6.27 Box and Whisker plot `RPupilSize`



6.5 Conclusion

Missing data is a universal problem in healthcare research. This chapter outlined varying approaches to management of missing data and detailed how missing data in the WMAS dataset were addressed. The next chapter is concerned with development of the SEPSIS screening tool and will describe the analysis and pooling phases of multiple imputation.

Chapter 7

Derivation of the SEPSIS screening tool

7.1 Introduction

This chapter describes how the **S**creening to **E**nhance **P**reho**S**pital **I**dentification of **S**epsis (SEPSIS) tool was developed. Development of the SEPSIS screening tool adheres to the methods proposed by Labarère *et al*,[58] previously described in **chapter 2**. Not all elements of the Labarère checklist are reported in this chapter. Validation of the SEPSIS screening tool is reported in **chapter 8**, while comparison of the SEPSIS screening tool with different models is reported in **chapter 9**.

7.2 Method

7.2.1 Rationale

It is estimated that 70% of sepsis cases arise in the community with two-thirds of severe sepsis cases initially seen in the emergency department (ED).[28] Half of all ED sepsis patients will arrive via emergency medical services (EMS).[26-28 170 171] Sepsis patients transported to the ED by EMS are likely to be sicker than those arriving by other means.[170-173] Sepsis can be very challenging to identify in the prehospital environment.[39] It has been suggested that a prehospital sepsis screening tool would help improve recognition of sepsis by ambulance clinicians.[40] Earlier recognition of sepsis may facilitate delivery of antibiotics, and other treatments, by ambulance clinicians, before arriving at hospital. Early intervention, before arrival at hospital, has the potential to improve outcomes for patients with sepsis.[174]

7.2.2 Objective

Develop a simple screening tool, using routinely collected patient demographic and clinical data, for use by ambulance clinicians, at the patient's bedside, to identify patients at high risk for sepsis (as per NICE clinical guideline 51). The NICE sepsis guideline stratifies patients into four risk categories: no risk (patients who do not have infection), low risk, moderate risk and high risk.

Patients classified as high risk should be treated with intravenous antibiotics without delay. Moderate risk and low risk patients do not require immediate antibiotic prophylaxis. If the objective of prehospital sepsis screening is to facilitate early intervention, it follows that any prehospital sepsis screening tool should identify those patients who require immediate antibiotic therapy as per the NICE guideline. Consequently, the objective of the study is to improve identification of patients with high risk of sepsis.

7.2.3 Study design

This study is a retrospective, cross-sectional study. The study utilises a retrospective dataset from West Midlands Ambulance Service NHS Foundation Trust (WMAS). The data include consecutive patients (cases) transported to the Emergency Department of University Hospital North Staffordshire NHS Foundation Trust (UHNS) between 01 July 2013 and 30 June 2014.

7.2.4 Participants and setting

The study setting is a UK NHS Ambulance Service. All patients attended by WMAS and transported to UHNS between 01 July 2013 and 30 June 2014 were eligible for inclusion. Exclusion criteria were age under 18 years, and all cases of cardiac arrest, trauma or mental health diagnosis. Only electronic patient report form (ePRF) records were available. Records stored on paper patient report forms (pPRF) was not intended to be an exclusion criteria, but WMAS were unable to obtain pPRF data due to technical challenges. The included population therefore comprised of adult patients, with a medical complaint, whose care was documented using the ePRF.

7.2.5 Outcome

The screening tool is intended to predict high risk sepsis as defined by the 2016 NICE guideline (NG51) "Sepsis: recognition, diagnosis and early management". Diagnostic criteria for sepsis are outlined in **table 7.1**.

Table 7.1 Classification of risk of sepsis (NICE 2016)

Category	High risk criteria	Moderate to high risk criteria	Low risk criteria
History	Objective evidence of new altered mental state	History from patient, friend or relative of new onset of altered behaviour or mental state History of acute deterioration of functional ability Impaired immune system (illness or drugs including oral steroids) Trauma, surgery or invasive procedures in the last 6 weeks	Normal behaviour
Respiratory	Raised respiratory rate: 25 breaths per minute or more New need for oxygen (more than 40% FiO2) to maintain saturation more than 92% (or more than 88% in known chronic obstructive pulmonary disease)	Raised respiratory rate: 21–24 breaths per minute	No high risk or moderate to high risk criteria met
Blood pressure	Systolic blood pressure 90 mmHg or less or systolic blood pressure more than 40 mmHg below normal	Systolic blood pressure 91–100 mmHg	No high risk or moderate to high risk criteria met
Circulation and hydration	Raised heart rate: more than 130 beats per minute Not passed urine in previous 18 hours. For catheterised patients, passed less than 0.5 ml/kg of urine per hour	Raised heart rate: 91–130 beats per minute (for pregnant women 100–130 beats per minute) or new onset arrhythmia Not passed urine in the past 12–18 hours For catheterised patients, passed 0.5–1 ml/kg of urine per hour	No high risk or moderate to high risk criteria met
Temperature		Tympanic temperature less than 36°C	
Skin	Mottled or ashen appearance Cyanosis of skin, lips or tongue Non-blanching rash of skin	Signs of potential infection, including redness, swelling or discharge at surgical site or breakdown of wound	No non-blanching rash

High risk sepsis status was determined by examination of UHNS clinical data, and not simply extracted from UHNS diagnostic fields. All records were assessed to determine if high risk sepsis criteria were present to ensure no patients were misclassified. Presence of any one high risk criteria results in a patient being classified as at high risk of sepsis.

7.2.6 Missing values

Missing data, and the management thereof, have previously been described in **Chapter 6**.

7.2.7 Candidate predictor variables

All patient demographic and clinical variables were eligible for inclusion in the clinical prediction model.

7.2.8 Sample size

The original data sample of 22945 cases was divided into derivation and validation cohorts. The derivation cohort is used to develop the SEPSIS screening tool, while the validation cohort is used to assess the performance of the SEPSIS screening tool (see **chapter 8**). It has been argued that, when developing a predictive model, at least ten instances of the outcome of interest (in this case high risk of sepsis) are required, per candidate predictor included in the model, to ensure statistically valid results.[59 69 71 175-177] Twenty-eight potential variables are available for inclusion in the model (see **table 6.5**). Consequently, at least 280 cases with high risk of sepsis must be included in the derivation cohort. Similarly, Vergouwe *et al*[178] argue that at least 100 events (high risk of sepsis) and 100 non-events (no sepsis) are required to assess model performance. However, Steyerberg[59] suggests that, to detect small differences in model performance, the validation sample should contain at least 250 cases of the outcome of interest (high risk of sepsis).

The R package `caret` (version 6.0-71) was used to divide the original dataset into derivation and validation cohorts, ensuring equal distribution of population characteristics. The dataset was divided to ensure 250 cases of high risk sepsis were available in the validation dataset. Ultimately the dataset was divided into 70% derivation, 30% validation cohorts. The derivation cohort comprised 13083 cases, while the validation cohort comprised 5607 cases. The number of cases classified as high risk of sepsis within each of the datasets is given in **table 7.2**.

Table 7.2 Sepsis cases in each imputed dataset

Dataset	No sepsis risk *n* (%)	Low Risk Sepsis *n* (%)	Moderate/High Risk Sepsis *n* (%)	High Risk Sepsis *n* (%)
WMAS (complete)	18690 (81.4)	1496 (6.5)	1912 (8.3)	847 (3.7)
WMAS (derivation)	13083 (81.4)	1048 (6.5)	1339 (8.3)	593 (3.7)
WMAS (validation)	5607 (81.5)	448 (6.5)	573 (8.3)	254 (3.7)

The derivation dataset contains 593 cases with high risk of sepsis, thus the maximum number of candidate predictor variables that could conceivably be considered during development of the screening tool is 59 (593/10, rounded down). The number of events per candidate predictor is 21 (593/28). The number of high risk of sepsis cases present in the derivation dataset is therefore sufficient to calculate reliable estimates. The validation dataset contains 254 cases of high risk of sepsis, sufficient to detect small changes in model performance.

7.2.9 Model specification

The outcome of interest is a binary outcome (presence or absence of high risk sepsis as per NICE Guideline 51), consequently logistic regression was required for model development. The outcome of interest is high risk of sepsis at hospital.

Simple logistic regression was undertaken to quantify the relationship between individual candidate predictor variables, and the outcome of interest. Regression coefficients are often used as a first stage to filter out variables from inclusion in the multivariate model. The threshold at which candidate predictor variables are excluded from the multivariable model clearly impacts the number of predictors selected. Only selecting highly correlated candidate predictor variables generates models with fewer predictors, with the trade-off of potentially missing useful predictors.[69] It can also result in over-fitting leading to optimistic models that yield biased model coefficient estimations.[59 69 71] To reduce the likelihood of overfitting, no candidate predictor variables were excluded following simple logistic regression.

Next, candidate predictor variables were assessed for multicollinearity. Multicollinearity occurs when a high correlation between two or more predictor variables and the outcome of interest is identified. Medical statisticians warn

against the use of explanatory variables that are not independent,[179-181] as use of highly correlated variables within regression models can give rise to spurious results.[182-184] In other words, if two predictor variables are highly correlated with each other, it may be appropriate to include only one of the variables in the prediction model. Within this study, highly correlated candidate predictor variables were reduced to a single predictor variable for inclusion in the multivariable model (see **section 3.2.2**).

Selection of independent predictor variables to be included in the parsimonious model was informed by stepwise selection. Stepwise selection of variables can be by either backward elimination or forward selection. Backward elimination is generally preferred to forward selection because the former assesses all candidate predictors simultaneously and performs better should candidate predictors be highly correlated with each other.[68] [71] In stepwise selection candidate predictors are removed from, or added to, the model in a sequential manner. The order in which variables are removed is dependent upon their correlation with the outcome of interest, expressed as their correlation coefficient. During backward elimination, the candidate predictor variable with the lowest correlation is removed from the full multivariable model first, then the second lowest correlated variable is removed and so on. Each time a variable is removed from the model, performance of 'the smaller model' is assessed relative to that of 'the larger model'.

Relative performance is assessed by calculating the Akaike Information Criteria (AIC). The AIC is an index that measures statistical goodness of fit relative to the number of variables included in a model. It imposes a penalty for increasing the number of variables. Lower values of the AIC indicate the preferred model, that is, it helps to identify the model with the fewest variables that provides an adequate fit to the data.[185] In other words, if the AIC calculated for 'the smaller model' (the model with the variable removed) is lower than the AIC for 'the larger model' (the model without the variable removed), it implies 'the smaller model' is a better model than 'the larger model'. The variable excluded from 'the smaller model' should not therefore be included in the parsimonious model. Conversely, if the AIC calculated for 'the larger model' is lower than the AIC of 'the smaller model', then 'the larger model' is the better model, and the omitted variable should be included in the parsimonious model.

Although AIC is useful to select the best model from a group of models, it does not provide a reliable estimate of model quality, that is, it is useful to select the best performing model, but the selected model will not necessarily be an accurate model.

In forward selection, the method is reversed. The first model comprises the candidate predictor variable with the greatest regression coefficient, and variables are added sequentially in order of decreasing regression coefficients. Comparison of models is similarly based upon estimation of the AIC.

An additional complexity arises due to the presence of missing data, and use of multiple imputation. The standard statistical approach to multiply imputed data has three distinct phases:

- Impute the missing data *m* times via Markov chained equations.
- Perform statistical analysis on each of the *m* imputed datasets independently.
- Pool the *m* statistical analyses calculated in phase 2.

Stepwise variable selection occurs during phase two, however because the *m* datasets are different, it can result in different selections of variables across the *m* datasets. In other words, backwards selection of variables occurs on each of the *m* datasets. However, because the data in each of the *m* datasets is different, it can result in different variables being selected across the *m* datasets. How then do we manage these disparate selections?

It is not immediately obvious how parameters should be pooled in phase three, should this occur. The approach adopted in this study adheres to recommendations by van Buuren,[127] who developed the `mice` package (used to impute data in this study). Put simply, the approach considers the number of times a variable is selected across the *m* datasets and utilises the Wald statistic to determine if the variable should be included in the multivariable model.

The Wald statistic quantifies the contribution of a variable (or group of variables) to the performance of a statistical model, and is estimated by maximum likelihood.[185] A significant Wald statistic ($p=<0.05$) indicates that the variable (or group of variables) impacts model performance, whereas a non-significant Wald

statistic ($p>0.05$) indicates that the variable (or group of variables) does not impact model performance.

For clarity, in the context of this screening tool development, the AIC is used to determine if a variable should be included in a model developed from a single dataset, whereas the Wald statistic is used to measure the contribution of a variable to model performance across multiple datasets.

Consequently, in the context of backward elimination processed on multiple imputed datasets, a significant Wald statistic indicates that the candidate predictor variable should be included in the parsimonious model. Conversely, a non-significant Wald statistic ($p>0.05$) indicates that the candidate predictor variable should be excluded from the parsimonious model. Wald statistics were calculated for all variables identified during stepwise selection to confirm if they should be included in the parsimonious model or not.

The final selection of independent predictor variables included in the parsimonious model was informed by the systematic review reported in **chapter 2**, stepwise selection as reported above, and practical utility.

7.2.10 Continuous predictors

Regression models with continuous predictor variables require computers or specialist software to generate answers. Consequently, regression models are unsuitable for use in the prehospital environment. It is usually necessary to simplify a model to enable bedside use, by the ambulance clinician, without requiring any additional equipment. A common approach to simplification is to reduce continuous variables into categorical variables, by subdividing the variable range into intervals.

Regression purists argue against transforming continuous variables into categorical variables, as to do so may lead to a loss of precision within the model.[59 71 175 186] However, loss of precision needs to be balanced against ease of use. Conversion of continuous predictor variables into categorical predictor variables using clinically meaningful thresholds from clinical practice that are easily remembered by clinicians is a common approach,[187 188] and has been adopted in this study.

Cut points for continuous predictors were determined pragmatically. Several approaches informed cut point selection. First, the normal physiologic range of clinical variables was considered. For example, a normal pulse rate would be in the range 60 -100 beats per minute. Second, cut points used in previously published screening tools were assessed. For example, the PreSep score identifies a cut point for respirations at 22 breaths per minute.[54] Third, Loess curves were plotted for each continuous predictor variable, with respect to risk of sepsis, to visually identify important thresholds. Fourth, theoretical cut points, to identify thresholds at which risk of sepsis increases or decreases, were calculated for each predictor variable. Cut points were calculated using the R package `OptimalCutpoints` (version 1.1-3).

`OptimalCupoints` allows the user to specify the method used to calculate upper and lower cut points. The method adopted in this thesis was *prevalence matching*. The upper cut point represents the upper boundary where sepsis becomes more common than the mean of the population. For example, when considering respiratory rate, the upper cut point is the boundary where the proportion of patients with sepsis is greater than 3.7% (the mean incidence of high risk of sepsis). Similarly, the lower cut point represents the boundary below which the incidence of sepsis is greater than the mean of the population.

When calculating cut points, several different diagnostic standards for sepsis were considered (NICE high, moderate and low of risk sepsis, SIRS sepsis, SOFA sepsis, and ED Doctor diagnosis of sepsis) to help provide greater detail.

Each of the above informed the decision of how to subdivide the continuous predictor variables into intervals for further assessment. To guard against the potential loss of precision, continuous predictor variables were subdivided into several small intervals, rather than a single threshold.

7.2.10.1 Model simplification

To enable bedside calculation of the SEPSIS screening tool, without the need for additional equipment, regression coefficients were converted into rounded scores to enable the creation of a simple arithmetic tool. Each regression coefficient was rounded to the nearest integer. The bedside clinician simply summates the score for all variables included in the model.

7.2.11 Model performance

7.2.11.1 Model calibration

The Hosmer-Lemeshow statistic was used to assess goodness of fit.[189] The Hosmer-Lemeshow test compares fitted values with observed values.[190] A statistically significant ($p<0.05$) Hosmer-Lemeshow statistic suggests that a model may be overfitted.[190] However, a significant Hosmer-Lemeshow statistic is often observed when models have been derived from large samples.[191] Where the Hosmer-Lemeshow statistic is significant, assessment of the model calibration slope provides an alternate means to determine if the model is overfitted.[192] The concordance statistic (c-statistic) is an alternate measure of goodness of fit and quantifies the predictive accuracy of a logistic regression model.[193] However, for a binary outcome, as in this case, the c-statistic is identical to area under the receiver operating characteristic (ROC) curve. [59 194] The ROC curve is calculated to assess model discrimination (see below), so there is no need to calculate the c-statistic.

7.2.11.2 Model discrimination

Model discrimination was assessed by calculating the area under the receiver operating characteristic curve (ROC) for the outcome of high sepsis risk at hospital. ROC curve analysis allows visual evaluation of the trade-offs between sensitivity and specificity associated with different values of the test result.[195] The ideal medical diagnostic model would have an ROC≥0.95, with good performance indicated by ROC≥0.80.[195]

7.2.11.3 Model operating characteristics

Model performance was assessed by calculating sensitivity, specificity, positive likelihood ratio, negative likelihood ratio and diagnostic odds ratio.

7.3 Results

7.3.1 Candidate predictor variables

The variables in **table 7.3** were excluded from consideration as candidate predictor variables. The majority were excluded as they were neither related to the patient or their clinical condition. Four potential candidate predictor variables were excluded due to very high missingness.

Table 7.3 Excluded candidate predictor variables

Variable	Reason for exclusion
PatientID	Unique ID. Not patient/clinically related
UHNS_ID	Unique ID. Not patient/clinically related
Dispatch Complaint1	Not patient/clinically related
Dispatch Complaint2	Not patient/clinically related
Dispatch Category	Not patient/clinically related
Forename	PID removed following linkage
Surname	PID removed following linkage
Address 1	PID removed following linkage
Address 2	PID removed following linkage
City	PID removed following linkage
County	PID removed following linkage
Post Code	PID removed following linkage
DOB	Age variable calculated in preference
Incident Address	PID removed following linkage
Incident City	PID removed following linkage
Incident Postcode	PID removed following linkage
Incident Time	Not patient/clinically related
Response Priority	Not patient/clinically related
Mobile	Not patient/clinically related
Arrive Scene	Not patient/clinically related
Leaving Scene	Not patient/clinically related
Destination Arrival	Not patient/clinically related
Peak Flow (l/min)	Very high missingness
End Tidal CO2 (kPa)	Very high missingness
Airway Signs	Very high missingness
Airway Status	Very high missingness
Arrest Occurance	Excluded case type
Crew Level 1	Not patient/clinically related
Crew Level 2	Not patient/clinically related
Institution Name	Not patient/clinically related

PID – personal identifiable data

All remaining variables were included as candidate predictor variables (see **table 7.4)**.

Table 7.4 Candidate predictor variables

Variable	Variable name	Type
Impression	Imp	Factor
Location	Location	Factor
Age	Age	Integer
Sex	Gender	Factor
Respiratory Rate (BPM)	Resps	Integer
Oxygen Saturation (%)	SpO2	Integer
Heart Rate (BPM)	Pulse	Integer
Systolic BP (mmHg)	SBP	Integer
Diastolic BP (mmHg)	DBP	Integer
Temperature (Celcius)	Temp	number
Blood Glucose (mmol/L)	BM	number
Glasgow Coma Scale	GCS_sum	Integer
GCS Eye	GCS_eye	Ordered Factor
GCS Verbal	GCS_verbal	Ordered Factor
GCS Motor	GCS_motor	Ordered Factor
Skin Colour	Skin	Factor
Cap Refill Time	CBRT	Ordered Factor
AVPU	AVPU	Ordered Factor
Pupil Right - Reactivity	RPupilReact	Ordered Factor
Pupil Right - Size	RPupilSize	Integer
Pupil Left - Reactivity	LPupilReact	Ordered Factor
Pupil Left - Size	LPupilSize	Integer
Oxygen	Oxygen	Factor
Fluids	Fluids	Factor

7.3.2 Model specification

7.3.2.1 Regression of candidate predictor variables

Regression coefficients, pooled following multiple imutation, for continuous predictor variables are reported in **table 7.5**, while those for categorical candidate predictors are reported in **table 7.6**.

Table 7.5 Regression of continuous candidate predictor variables

Variable	intercept	ß	95% CI	p-value
Age	-5.2	0.028	0.023 to 0.033	<0.001
Resps	-5.9	0.11	0.10 to 0.12	<0.001
SpO2	6.6	-0.11	-0.11 to -0.96	<0.001
Pulse	-5.8	0.026	0.023 to 0.028	<0.001
SBP	-1.4	-0.015	-0.018 to -0.011	<0.001
DBP	-1.9	-0.018	-0.023 to -0.013	<0.001
Temp	-33.1	0.8	0.72 to 0.88	<0.001
BM	-3.6	0.053	0.032 to 0.074	<0.001
GCS_sum	-1.2	-0.15	-0.18 to -0.12	<0.001
RPupilSize	-2.9	-0.12	-0.23 to -0.0089	0.034
LPupilSize	-2.9	-0.12	-0.23 to -0.0057	0.039

ß – regression coefficient, CI – confidence interval

Table 7.6 Regression of categorical candidate predictor variables

Variable	Category	intercept	$ß_i$	95% CI	p-value
Imp	General medical		Reference category		
	Other	-3.5	-0.45	-0.26 to 1.20	0.21
	Cardiovascular	-3.5	-0.36	-0.68 to -0.044	0.026
	Neurological	-3.5	-1.80	-2.50 to -1.20	<0.001
	Obstetric/Gynae	-3.5	-13.00	-340 to 310	0.94
	Respiratory	-3.5	1.50	1.30 to 1.70	<0.001
Location	Home		Reference category		
	Nursing home	-3.3	1.20	0.95 to 1.40	<0.001
	Other	-3.3	-0.94	-1.20 to -0.65	<0.001
Gender	Male		Reference category		
	Female	-3.2	-0.026	-0.19 to 0.14	0.75
GCS_eye	Spontaneous		Reference category		
	Verbal	-3.4	1.00	0.74 to 1.30	<0.001
	Pain	-3.4	1.10	0.58 to 1.50	<0.001
	No response	-3.4	1.10	0.60 to 1.60	<0.001
GCS_verbal	Oriented		Reference category		
	Confused	-3.4	0.44	0.18 to 0.69	<0.001
	Inappropriate words	-3.4	1.40	0.86 to 1.90	<0.001
	Incomprehensible sounds	-3.4	1.30	0.87 to 1.70	<0.001

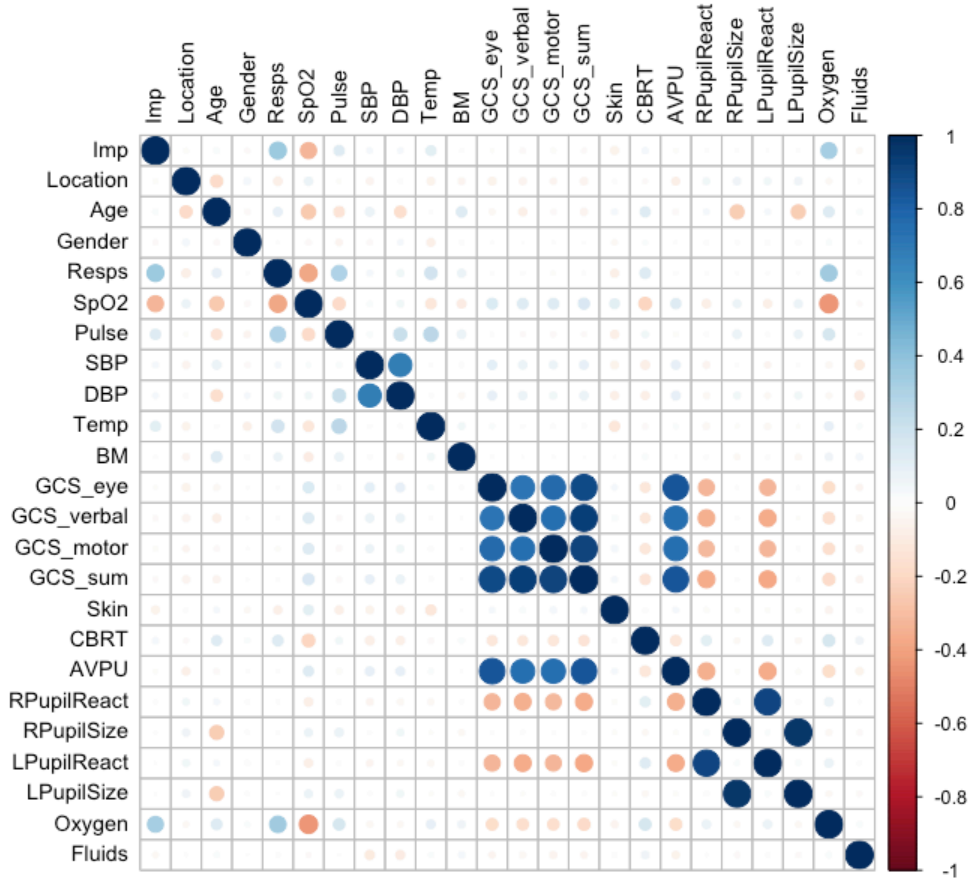
	No response	-3.4	1.30	0.91 to 1.60	<0.001
GCS_motor	Obeys	Reference category			
	Localises	-3.4	1.20	0.87 to 1.50	<0.001
	Withdraws	-3.4	1.20	0.80 to 1.60	<0.001
	Flexion	-3.4	0.83	-0.61 to 2.30	0.26
	Extension	-3.4	2.30	0.98 to 3.60	<0.001
	No response	-3.4	1.20	0.72 to 1.70	<0.001
Skin	Normal	Reference category			
	Cyanosed	-3.7	1.80	1.40 to 2.30	<0.001
	Flushed	-3.7	1.10	0.82 to 1.40	<0.001
	Jaundice	-3.7	1.50	0.82 to 2.20	<0.001
	Mottling	-3.7	2.20	1.40 to 2.90	<0.001
	Pallor	-3.7	1.10	0.95 to 1.30	<0.001
	Rash	-3.7	-11.0	-390 to 370	0.96
CBRT	Normal	Reference category			
	Delayed	-3.4	1.30	1.00 to 1.50	<0.001
AVPU	Alert	Reference category			
	Verbal	-3.4	0.96	0.70 to 1.20	0.63
	Pain	-3.4	1.10	0.75 to 1.50	0.90
	No response	-3.4	0.99	0.49 to 1.50	0.001
RPuPilReact	Brisk	Reference category			
	Sluggish	-3.3	0.55	0.26 to 0.84	<0.001
	Fixed	-3.3	-0.40	-1.60 to 0.74	0.49
LPuPilReact	Brisk	Reference category			
	Sluggish	-3.3	0.59	0.31 to 0.88	<0.001
	Fixed	-3.3	0.28	-0.62 to 1.20	0.54
Oxygen	No oxygen given	Reference category			
	Oxygen given	-3.7	1.9	1.70 to 2.00	<0.001
Fluids	No fluids given	Reference category			
	Fluids given	-3.3	1.9	1.40 to 2.40	<0.001

ß – regression coefficient, CI – confidence interval

7.3.2.2 Test for multicollinearity between candidate predictor variables

Multicollinearity was assessed by identifying correlations among individual candidate predictor variables. **Figure 7.1** graphically displays the relationship between variables. Correlation coefficients between candidate predictor variables are reported in **Table 7.7**

Figure 7.1. Correlations between candidate predictor variables



Blue circles indicate a positive relationship, while red circles indicate an inverse relationship. Darker colour shades and increasing circle size indicate increasing strength of the relationship. The dark blue diagonal represents the perfect relationship of each variable with itself.

A near perfect positive relationship is seen between left and right pupils with respect to reactions and pupil size, that is `LPupilReact` = `RPupilReact` and `LPupilSize` ≈ `RPupilSize`. It can also be seen that there is a very strong correlation between `GCS_sum` and the components of the Glasgow Coma Score (`GCS_eye`, `GCS_verbal` and `GCS_motor`). This is to be expected as `GCS_sum` is a composite score of the Glasgow Coma Score components. Similarly, there is a positive relationship between `AVPU` and the Glasgow Coma score (sum and

components). This too is to be expected as both Glasgow Coma Score and AVPU are both measures of consciousness. An additional positive relationship is apparent between systolic and diastolic blood pressures.

Fewer inverse relationships are visible, and all are considerably weaker. The most notable inverse relationships occur between oxygen use (`oxygen`) and peripheral oxygen saturations (SpO_2). This is because patients with reduced oxygen saturations will be treated with oxygen. Similarly, there appears to be a relationship between measures of consciousness (`AVPU`, `GCS_sum`, `GCS_eye`, `GCS_verbal` and `GCS_motor`) and pupil reactions (`LPupilReact` and `RPupilReact`) which is consistent with pupil reactions becoming increasingly sluggish or slow when consciousness is considerably reduced.

Table 7.7 Correlations between variables

	Imp	Location	Age	Gender	Resps	SpO_2	Pulse	SBP	DBP	Temp	BM	GCS_eye	GCS_verbal	GCS_motor	GCS_sum	Skin	CBRT	AVPU	RPupilReact	RPupilSize	LPupilReact	LPupilSize	Oxygen	Fluids
Imp	1.00	-0.02	0.03	-0.02	0.40	-0.30	0.10	0.03	0.03	0.10	0.01	-0.01	-0.03	-0.02	-0.02	-0.07	0.04	-0.01	<0.01	<0.01	<0.01	<0.01	0.30	-0.03
Location	-0.02	1.00	-0.20	0.05	-0.08	0.08	-0.01	-0.05	0.01	-0.06	-0.05	-0.07	-0.06	-0.04	-0.06	0.01	-0.02	-0.07	0.05	0.06	0.05	0.06	-0.03	-0.01
Age	0.03	-0.20	1.00	-0.03	0.10	-0.30	-0.10	0.09	-0.20	0.01	0.10	-0.04	-0.08	-0.03	-0.05	0.03	0.10	-0.03	0.04	-0.20	0.03	-0.20	0.10	0.03
Gender	-0.02	0.05	-0.03	1.00	<0.01	-0.03	-0.05	-0.03	0.03	-0.07	0.02	<0.01	<0.01	<0.01	<0.01	-0.03	0.02	<0.01	0.02	0.01	0.02	0.01	0.02	<0.01
Resps	0.40	-0.08	0.10	<0.01	1.00	-0.40	0.30	0.04	0.06	0.20	0.07	<0.01	-0.01	-0.01	-0.01	-0.07	0.10	0.01	-0.01	<0.01	-0.01	<0.01	0.30	0.01
SpO_2	-0.30	0.08	-0.30	-0.03	-0.40	1.00	-0.20	0.02	0.05	-0.10	-0.09	0.10	0.10	0.10	0.20	0.10	-0.20	0.10	-0.08	0.07	-0.08	0.07	-0.40	-0.02
Pulse	0.10	-0.01	-0.10	-0.05	0.30	-0.20	1.00	0.03	0.20	0.30	0.08	-0.01	-0.02	-0.03	-0.02	-0.08	0.05	-0.02	0.01	0.08	0.01	0.08	0.20	<0.01
SBP	0.03	-0.05	0.09	-0.03	0.04	0.02	0.03	1.00	0.70	0.01	0.01	0.10	0.08	0.07	0.09	-0.06	-0.09	0.09	-0.05	-0.01	-0.05	-0.01	-0.04	-0.10
DBP	0.03	0.01	-0.20	0.03	0.06	0.05	0.20	0.70	1.00	-0.01	-0.04	0.09	0.08	0.06	0.08	-0.07	-0.08	0.09	-0.04	0.05	-0.03	0.05	-0.03	-0.10
Temp	0.10	-0.06	0.01	-0.07	0.20	-0.10	0.30	0.01	-0.01	1.00	0.05	0.02	0.01	0.02	0.01	-0.10	-0.03	0.02	-0.04	-0.01	-0.04	-0.01	0.09	0.02
BM	0.01	-0.05	0.10	0.02	0.07	-0.09	0.08	0.01	-0.04	0.05	1.00	0.01	<0.01	<0.01	<0.01	<0.01	0.03	<0.01	0.01	-0.03	0.01	-0.03	0.06	0.04
GCS_eye	-0.01	-0.07	-0.04	<0.01	<0.01	0.10	-0.01	0.10	0.09	0.02	0.01	1.00	0.70	0.80	0.90	0.03	-0.10	0.80	-0.30	-0.01	-0.30	-0.01	-0.20	-0.05
GCS_verbal	-0.03	-0.06	-0.08	<0.01	-0.01	0.10	-0.02	0.08	0.08	0.01	<0.01	0.70	1.00	0.70	0.90	0.03	-0.10	0.70	-0.40	-0.01	-0.40	-0.01	-0.20	-0.04
GCS_motor	-0.02	-0.04	-0.03	<0.01	-0.01	0.10	-0.03	0.07	0.06	0.02	<0.01	0.80	0.70	1.00	0.90	0.04	-0.10	0.80	-0.30	-0.01	-0.30	-0.01	-0.20	-0.05
GCS_sum	-0.02	-0.06	-0.05	<0.01	-0.01	0.20	-0.02	0.09	0.08	0.01	<0.01	0.90	0.90	0.90	1.00	0.04	-0.10	0.80	-0.40	-0.01	-0.40	-0.01	-0.20	-0.05
Skin	-0.07	0.01	0.03	-0.03	-0.07	0.10	-0.08	-0.06	-0.07	-0.10	<0.01	0.03	0.03	0.04	0.04	1.00	0.01	0.03	-0.02	-0.01	-0.02	-0.01	-0.05	0.02
CBRT	0.04	-0.02	0.10	0.02	0.10	-0.20	0.05	-0.09	-0.08	-0.03	0.03	-0.10	-0.10	-0.10	-0.10	0.01	1.00	-0.10	0.10	-0.03	0.10	-0.03	0.20	0.06
AVPU	-0.01	-0.07	-0.03	<0.01	0.01	0.10	-0.02	0.09	0.09	0.02	<0.01	0.80	0.70	0.80	0.80	0.03	-0.10	1.00	-0.40	-0.01	-0.40	-0.02	-0.20	-0.06
RPupilReact	<0.01	0.05	0.04	0.02	-0.01	-0.08	0.01	-0.05	-0.04	-0.04	0.01	-0.30	-0.40	-0.30	-0.40	-0.02	0.10	-0.40	1.00	0.03	0.90	0.02	0.08	<0.01
RPupilSize	<0.01	0.06	-0.20	0.01	<0.01	0.07	0.08	-0.01	0.05	-0.01	-0.03	-0.01	-0.01	-0.01	-0.01	-0.01	-0.03	-0.01	0.03	1.00	0.02	1.00	-0.03	-0.03
LPupilReact	<0.01	0.05	0.03	0.02	-0.01	-0.08	0.01	-0.05	-0.03	-0.04	0.01	-0.30	-0.40	-0.30	-0.40	-0.02	0.10	-0.40	0.90	0.02	1.00	0.03	0.08	0.01
LPupilSize	<0.01	0.06	-0.20	0.01	<0.01	0.07	0.08	-0.01	0.05	-0.01	-0.03	-0.01	-0.01	-0.01	-0.01	-0.01	-0.03	-0.02	0.02	1.00	0.03	1.00	-0.03	-0.03
Oxygen	0.30	-0.03	0.10	0.02	0.30	-0.40	0.20	-0.04	-0.03	0.09	0.06	-0.20	-0.20	-0.20	-0.20	-0.05	0.20	-0.20	0.08	-0.03	0.08	-0.03	1.00	0.03
Fluids	-0.03	-0.01	0.03	<0.01	0.01	-0.02	<0.01	-0.10	-0.10	0.02	0.04	-0.05	-0.04	-0.05	-0.05	0.02	0.06	-0.06	<0.01	-0.03	0.01	-0.03	0.03	1.00

Perfect correlation occurs when the correlation coefficient is 1. This can be 1 (a positive correlation) or -1 (a negative/inverse correlation). There is no established method to manage predictor variables that are highly correlated. The threshold at which a candidate predictor variable is deemed to be highly correlated is subjective. In this study, a correlation coefficient of 0.9 (positive or negative) was the threshold to assess the potential impact of highly correlated candidate predictor variables on the model.

A perfect correlation exists between left and right pupil reactions (correlation coefficient =1). Pupil size between left and right eyes are also very highly correlated (correlation coefficient =0.9). Differences between a patient's pupil reactions, and pupil sizes, are known to be associated with cerebral pathologies, however no such link has been established for sepsis. Furthermore, the systematic review reported earlier in this thesis failed to identify any known link between pupil size or reaction and sepsis. Therefore, analysing data for both pupils will not be of benefit over analysing data from one pupil. Consequently, data relating to the left pupil were excluded from further modelling, so that only data pertaining to the right pupil were employed during modelling.

The strong relationships between elements of the Glasgow Coma Score, the summated Glasgow Coma Score and AVPU score were more challenging to manage. Existing scores or screening tools utilise different measures of consciousness. For example, the Critical Illness Score utilises Glasgow Coma Score while the NEWS score utilises AVPU. It was not clear at this early stage which measure of consciousness, if any, would be the optimal measure for inclusion in the prediction model being developed. Consequently, each measure of consciousness (Glasgow Coma Score elements, the summated Glasgow Coma Score and AVPU) was modelled separately. That is, one model was generated using the Glasgow Coma Score elements as the measure of consciousness (Model A), another was generated using the Glasgow Coma Score sum as the measure of consciousness (Model B) and a third was generated using the AVPU score as the measure of consciousness (Model C). Performance of the three models (using differing measures of consciousness) were subsequently compared to determine the best performing model.

7.3.2.3 Generation of multivariable models

The following variables were previously identified as being used in existing screening tools during the systematic review (see **Chapter 3**, **table 3.6**):

- Respiratory rate
- Heart rate
- Temperature
- Level of consciousness
- Peripheral oxygen saturations
- Blood pressure
- Shock Index
- Lactate
- Blood glucose
- Skin
- Capillary bed refill time
- Dispatch category
- Location
- Age

Dispatch category is not considered as a potential predictor in this study as it is not directly patient related. Dispatch category is assigned during the initial emergency call and is related to the call priority assigned by the call taker. Furthermore, there are numerous dispatch categorisation models in use internationally, consequently inclusion of dispatch categorisation would limit the generalisability of this screening tool if it was included. Lactate is not considered in this screening tool as it is not measured by ambulance clinicians in the UK.

Allgöwer and Buri proposed the shock index in 1967 to quantify the degree of hypovolaemia in haemorrhagic and infectious shock states.[196] Shock index is a bedside assessment defined as heart rate divided by systolic blood pressure. The normal range is 0.5 to 0.7 in healthy adults. Shock index is not considered as a potential predictor in this study as it requires calculation, adding additional complexity for the attending ambulance clinician, and may result in erroneous conclusions in the event of mis-calculation.

Stepwise selection was applied to the derivation dataset to identify predictor variables. A backward elimination procedure was employed to identify variables

for consideration of inclusion in the parsimonious model. Three parallel models were developed using different measures of consciousness. Model A utilised the Glasgow Coma Score (GCS_sum), Model B utilised the components of the Glasgow Coma Score (GCS_eye, GCS_verbal and GCS_motor) while model C utilised the AVPU score. Results of the stepwise selection are reported in **table 7.8**.

Table 7.8 Selection of predictor variables

Model A	Number of selections	Model B	Number of selections	Model C	Number of selections
Imp	50	Imp	50	Imp	50
Location	50	Location	50	Location	50
Age	50	Age	50	Age	50
Gender	0	Gender	0	Gender	0
Resps	50	Resps	50	Resps	50
SpO2	50	SpO2	50	SpO2	50
Pulse	50	Pulse	50	Pulse	50
SBP	50	SBP	50	SBP	50
DBP	0	DBP	0	DBP	0
Temp	50	Temp	50	Temp	50
BM	1	BM	1	BM	0
Skin	50	Skin	50	Skin	50
CBRT	0	CBRT	0	CBRT	0
RPupilReact	11	RPupilReact	12	RPupilReact	8
RPupilSize	0	RPupilSize	0	RPupilSize	0
Oxygen	50	Oxygen	50	Oxygen	50
Fluids	50	Fluids	50	Fluids	50
GCS_sum	50	GCS_eye	0	AVPU	50
		GCS_verbal	50		
		GCS_motor	0		

To determine which of the variables, identified during stepwise selection, should be included in the parsimonious model, Wald statistics were calculated on the pooled results, and are reported in **table 7.9**.

Table 7.9 Wald statistics for variables identified during stepwise selection

Model A	Wald statistic	Model B	Wald statistic	Model C	Wald statistic
Imp	<0.001	Imp	<0.001	Imp	<0.001
Location	0.002	Location	0.004	Location	0.003
Age	<0.001	Age	<0.001	Age	<0.001
Resps	<0.001	Resps	<0.001	Resps	<0.001
SpO2	<0.001	SpO2	<0.001	SpO2	<0.001
Pulse	0	Pulse	0	Pulse	0
SBP	<0.001	SBP	<0.001	SBP	<0.001
Temp	0	Temp	0	Temp	0
BM	0.86	BM	0.85	BM	0.87
Skin	0.007	Skin	0.007	Skin	0.014
RPupilReact	0.42	RPupilReact	0.41	RPupilReact	0.46
Oxygen	<0.001	Oxygen	<0.001	Oxygen	<0.001
Fluids	<0.001	Fluids	<0.001	Fluids	<0.001
GCS_sum	<0.001	GCS_verbal	<0.001	AVPU	<0.001

The variables identified by stepwise selection are largely consistent with those utilised in existing screening tools. Two of the variables identified by stepwise selection had non-significant Wald statistics (`BM` and `RPupilReact`) and are thus excluded from further consideration. Of the remainder, clinical impression (`Imp`), oxygen administration (`Oxygen`) and intravascular fluid therapy (`Fluids`) are not used in existing sepsis screening tools.

`Imp` is a subjective assessment made by the attending ambulance crew that categorises the patients underlying condition as being either “general medical”, “other”, “cardiovascular”, “neurologic”, “obstetric or gynaecologic” or “respiratory” in nature. It is unclear what the inter-rater reliability (between ambulance clinicians) would be for this variable. For example, two different clinicians might classify a case of suspected pulmonary embolus as “respiratory” due to respiratory system involvement, “cardiovascular” due vascular system involvement or “other” due to the occlusive nature of the underlying condition. Due to the subjective nature of this variable, and a lack of guidance from the ambulance service in how to categorise patients, a pragmatic decision was made to exclude `Imp` from further consideration.

Although both oxygen administration (Oxygen) and Fluid administration (Fluids) were identified as independent predictors of high risk of sepsis, it is important to note that both are interventions performed by ambulance clinicians, rather than observable clinical characteristics derived from the patient. Furthermore, administration of oxygen (Oxygen) is triggered when peripheral oxygen saturation (SpO_2) falls below 93%. Similarly, intravenous fluids should only be administered (Fluids) when the systolic blood pressure (SBP) falls below 90mmHg, and then only by paramedic level staff. Because both peripheral oxygen saturations (SpO_2) and systolic blood pressure (SBP) are measured, and identified as independent predictors for high risk of sepsis, a pragmatic decision was made to exclude both oxygen administration (Oxygen) and fluid administration (Fluids) from further consideration. The variables ultimately included as independent predictor variables are reported in **table 7.10**.

Table 7.10 Final selection of independent predictor variables

Model A	Wald statistic	Model B	Wald statistic	Model C	Wald statistic
Location	0.002	Location	0.004	Location	0.003
Age	<0.001	Age	<0.001	Age	<0.001
Resps	<0.001	Resps	<0.001	Resps	<0.001
SpO_2	<0.001	SpO_2	<0.001	SpO_2	<0.001
Pulse	<0.001	Pulse	<0.001	Pulse	<0.001
SBP	<0.001	SBP	<0.001	SBP	<0.001
Temp	<0.001	Temp	<0.001	Temp	<0.001
Skin	0.007	Skin	0.007	Skin	0.014
GCS_sum	<0.001	GCS_verbal	<0.001	AVPU	<0.001

7.3.3 Continuous predictors

The ranges of continuous independent predictor variables are subdivided into small manageable intervals to minimise the impact of conversion into categorical variables and subsequently to enable generation of scores that facilitate calculation of a simple summative risk score. In all cases a reference interval is established and all other intervals are subject to logistic regression against the reference level.

Age

It is generally accepted that the risk of sepsis increases with advancing age,[3] however there are no established thresholds for age, above which, one can state with conviction that the risk of sepsis has increased. **Table 7.11** identifies thresholds previously established in the literature, while **table 7.12** reports calculated cut points calculated with the package `OptimalCutpoints` (version 1.1-3) for several sepsis diagnoses with respect to age. It can be seen from **tables 7.11** and **7.12** that there is considerable variation in thresholds.

Table 7.11 Existing thresholds for age

	range
Normal physiologic range	-
NICE Sepsis (low risk)	-
NICE Sepsis (mod risk)	-
NICE Sepsis (high risk)	-
NEWS (0)	-
NEWS (1)	-
NEWS (2)	-
NEWS (3)	-
UK Sepsis Trust (red flag)	-
UK Sepsis Trust (amber flag)	-
Critical Illness Score (0)	<45
Critical Illness Score (1)	>44
Critical Illness Score (2)	-
PRESS (0)	18 – 39
PRESS (1)	-
PRESS (2)	>60
PRESS (3)	-
PRESS (4)	40 - 59
PRESS (5)	-
PreSep	-

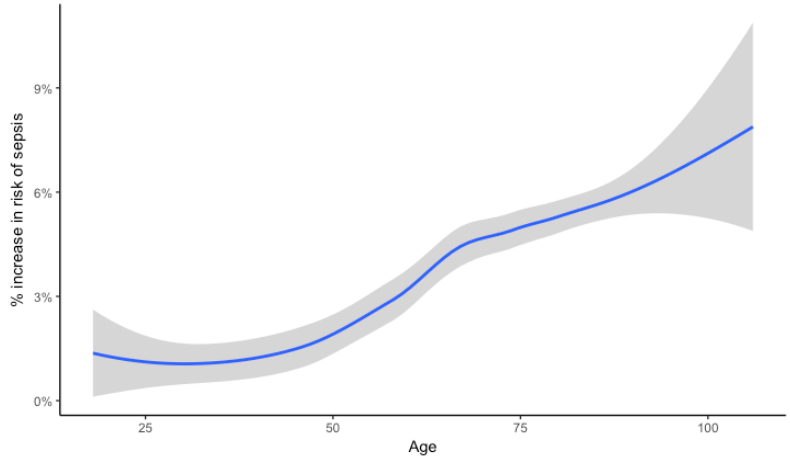
Table 7.12 Cut points for Age

Sepsis diagnosis		lower cut point	upper cut point
ED Doctor diagnosis	estimate	20	94
	AUC (95%CI)	*0.352 (0.326-0.378)*	*0.648 (0.622-0.674)*
Infection + 2SIRS	estimate	29	88
	AUC (95%CI)	*0.359 (0.346-0.372)*	*0.641 (0.628-0.654)*
SSC SOFA	estimate	20	94
	AUC (95%CI)	*0.301 (0.279-0.324)*	*0.699 (0.676-0.721)*
NICE (low, mod, high risk)	estimate	39	84
	AUC (95%CI)	*0.357 (0.346-0.367)*	*0.643 (0.633-0.654)*
NICE (mod & high risk)	estimate	31	87
	AUC (95%CI)	*0.378 (0.365-0.391)*	*0.622 (0.609-0.635)*
NICE (high risk only)	estimate	21	92
	AUC (95%CI)	*0.357 (0.337-0.378)*	*0.643 (0.622-0.663)*

ED – Emergency Department, SIRS – Systemic Inflammatory Response Criteria, SSC – Survive Sepsis Campaign, SOFA – Sequential Organ Failure Score, AUC – area under the curve

Figure 7.2 graphically indicates that below 40 to 50 years of age there is minimal change in risk of 'high risk sepsis'. At a little below 50 years of age risk of sepsis begins to increase at a more-or-less steady rate. Consequently, age below 40 years was adopted as a reference standard, with increments established every ten years.

Figure 7.2 Loess curve for Age vs NICE high sepsis risk



Respiratory rate

The normal physiological range for respirations is 10 to 20 breaths per minute.[197] **Table 7.13** identifies thresholds previously established in the literature, while **table 7.14** reports theoretical cut points for different sepsis diagnoses with respect to respiratory rate. It appears that respiratory rates below 10 to 12 breaths per minute are established as a lower threshold in the literature, while calculated lower thresholds range between 10 and 15 breaths per minute. The upper threshold varies considerably between 20 and 40 breaths per minutes.

Table 7.13 Existing thresholds for respirations

	range
Normal physiologic range	10 - 20
NICE Sepsis (low risk)	-
NICE Sepsis (mod risk)	21 - 24
NICE Sepsis (high risk)	>25
NEWS (0)	12-20
NEWS (1)	9 – 11
NEWS (2)	21-24
NEWS (3)	<9 or >24
UK Sepsis Trust (red flag)	>24
UK Sepsis Trust (amber flag)	21-24
Critical Illness Score (0)	12 - 23
Critical Illness Score (1)	<12 or 24 - 35
Critical Illness Score (2)	>35
PRESS (0)	-
PRESS (1)	-
PRESS (2)	-
PRESS (3)	-
PRESS (4)	-
PRESS (5)	-
PreSep	>22

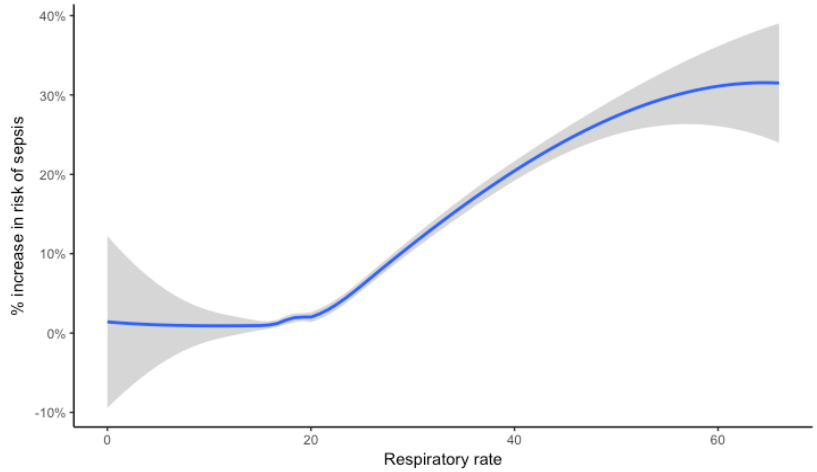
Table 7.14 Cut points for Respirations

Sepsis diagnosis		lower cut point	upper cut point
ED Doctor diagnosis	estimate	12	40
	AUC (95%CI)	*0.371 (0.34-0.402)*	*0.629 (0.598-0.66)*
Infection + 2SIRS	estimate	15	29
	AUC (95%CI)	*0.319 (0.304-0.333)*	*0.681 (0.667-0.696)*
SSC SOFA	estimate	12	40
	AUC (95%CI)	*0.36 (0.33-0.39)*	*0.64 (0.61-0.67)*
NICE (low, mod, high risk)	estimate	16	24
	AUC (95%CI)	*0.379 (0.368-0.391)*	*0.621 (0.609-0.632)*
NICE (mod & high risk)	estimate	15	28
	AUC (95%CI)	*0.325 (0.312-0.339)*	*0.675 (0.661-0.688)*
NICE (high risk only)	estimate	14	38
	AUC (95%CI)	*0.218 (0.197-0.239)*	*0.782 (0761-0.803)*

ED – Emergency Department, SIRS – Systemic Inflammatory Response Criteria, SSC – Survive Sepsis Campaign, SOFA – Sequential Organ Failure Score, AUC – area under the curve

Figure 7.3 graphically indicates that risk of 'high risk sepsis' is unchanging below 18 breaths per minute, but begins to increase at a rapid rate at approximately 20 breaths per minute.

Figure 7.3 Loess curve for Respirations vs NICE high sepsis risk



Because the normal physiological range begins at 10 breaths per minute this was taken to be the lower threshold. Similarly, the upper threshold was set at 20 breaths per minute as this is approximately the point when risk of sepsis begins to increase. The reference range is thus 10 to 20 breaths per minute. Increments

of 5 breaths per minute were established given the wide variation in the published literature.

Peripheral oxygen saturations

The normal physiological range for peripheral oxygen saturations is above 93% in patients who do not have chronic obstructive pulmonary disease.[198] **Table 7.15** identifies thresholds previously established in the literature, while **table 7.16** reports calculated cut points for different sepsis diagnoses with respect to peripheral oxygen saturations. It appears that peripheral oxygen saturations below 93% are established as a lower threshold in the literature, while calculated cut points range between 80% and 95%.

Table 7.15 Existing thresholds for peripheral oxygen saturations

	range
Normal physiological range	>94
NICE Sepsis (low risk)	-
NICE Sepsis (mod risk)	-
NICE Sepsis (high risk)	<93
NEWS (0)	>95
NEWS (1)	94 - 95
NEWS (2)	92 - 93
NEWS (3)	<92
UK Sepsis Trust (red flag)	<93
UK Sepsis Trust (amber flag)	-
Critical Illness Score (0)	>91
Critical Illness Score (1)	<92
Critical Illness Score (2)	-
PRESS (0)	>89
PRESS (1)	80 – 89
PRESS (2)	-
PRESS (3)	70 – 79
PRESS (4)	60 - 69
PRESS (5)	<60
PreSep	<92

Table 7.16 Cut points for Oxygen saturations (SpO2)

Sepsis diagnosis		**lower cut point**	**upper cut point**
ED Doctor diagnosis	estimate	79	100
	AUC (95%CI)	*0.663 (0.634-0.693)*	*0.337 (0.307-0.366)*
Infection + 2SIRS	estimate	91	100
	AUC (95%CI)	*0.71 (0.696-0.724)*	*0.29 (0.276-0.304)*
SSC SOFA	estimate	79	100
	AUC (95%CI)	*0.675 (0.646-0.703)*	*0.325 (0.297-0.354)*
NICE (low, mod, high risk)	estimate	94	99
	AUC (95%CI)	*0.657 (0.646-0.668)*	*0.343 (0.332-0.354)*
NICE (mod & high risk)	estimate	92	100
	AUC (95%CI)	*0.692 (0.679-0.706)*	*0.308 (0.294-0.321)*
NICE (high risk only)	estimate	84	100
	AUC (95%CI)	*0.775 (0.754-0.797)*	*0.225 (0.203-0.246)*

ED – Emergency Department, SIRS – Systemic Inflammatory Response Criteria, SSC – Survive Sepsis Campaign, SOFA – Sequential Organ Failure Score, AUC – area under the curve

Figure 7.4 graphically indicates that risk of sepsis increases below 95% (approximately), risk increases at a rapid rate until peripheral oxygen saturations approach 60%, after which risk appears to reduce once more. Peripheral oxygen saturations are a non-invasive estimate of arterial oxyhaemoglobin saturations, which in turn are dependent upon the partial pressure of oxygen in the arterial blood stream. The relationship between partial pressure of oxygen and oxyhaemoglobin saturations is described by the oxyhaemoglobin dissociation curve (see **figure 7.5)**. The oxyhaemoglobin dissociation curve is sigmoid in nature, and while the partial pressure of oxygen is above 80 mmHg, peripheral oxygen saturations remain above 95%. However, relatively small decreases in partial pressure of oxygen, below about 75mmHg, can produce significant reductions in peripheral oxygen saturations. For this reason, peripheral oxygen saturation readings below 75% can be prone to error. Furthermore, peripheral oxygen saturation readings are also prone to error whenever perfusion of the finger nail bed, where the measuring sensor is normally applied, is poor. Inadequate perfusion is a hallmark feature of patients

with sepsis, thus inaccurate peripheral oxygen saturation readings are not unexpected among patients with sepsis. Peripheral oxygen saturations below 75% may be unreliable.

Figure 7.4 Loess curve for peripheral oxygen saturations (SpO_2) vs NICE high sepsis risk

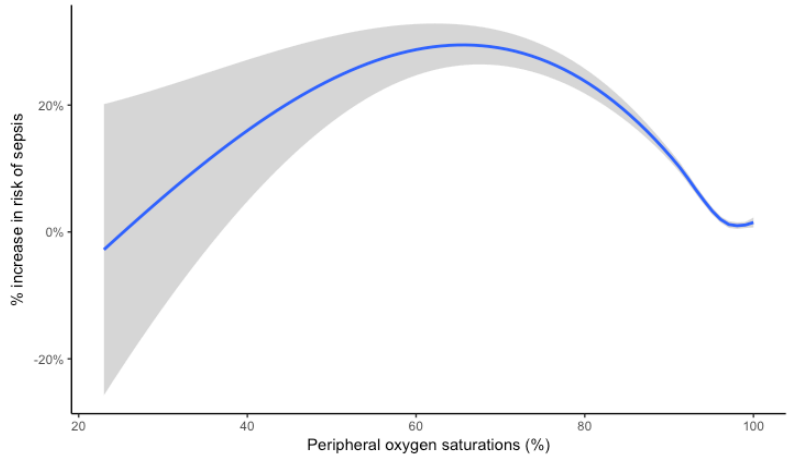


Figure 7.5 Oxygen dissociation curve

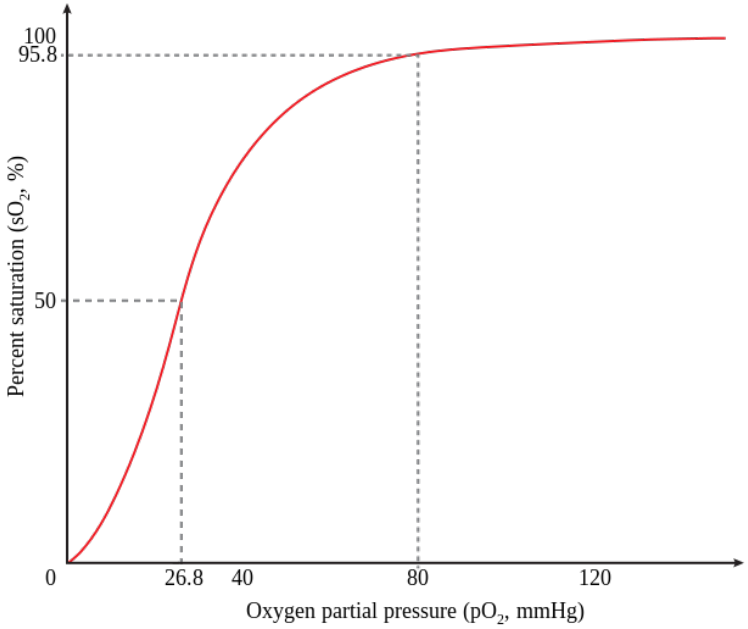


image from: https://en.wikipedia.org/wiki/Hemoglobin#/media/File:Hemoglobin_saturation_curve.svg

Because the normal physiological range for peripheral oxygen saturations is above 93%, this was taken to be the reference threshold. Further intervals were established in the ranges 89% to 93%, 85% to 88% and below 85%.

Heart rate

A heart rate below 60 beats per minute is established in the literature as the onset of bradycardia, while heart rates above 100 beats per minute are categorised as tachycardia.[197] The widely accepted normal physiological range for heart rate is therefore 60 to 100 beats per minute.[197] **Table 7.17** identifies thresholds previously established in the literature, while **table 7.18** reports calculated cut points for different sepsis diagnoses with respect to heart rate.

Table 7.17 Existing thresholds for heart rate

	range
Normal physiological range	60 - 100
NICE Sepsis (low risk)	-
NICE Sepsis (mod risk)	91-130
NICE Sepsis (high risk)	>129
NEWS (0)	51 - 90
NEWS (1)	41-50 or 91-110
NEWS (2)	111 - 130
NEWS (3)	<41 or >130
UK Sepsis Trust (red flag)	>130
UK Sepsis Trust (amber flag)	111 - 130
Critical Illness Score (0)	<121
Critical Illness Score (1)	>120
Critical Illness Score (2)	-
PRESS (0)	-
PRESS (1)	-
PRESS (2)	-
PRESS (3)	-
PRESS (4)	-
PRESS (5)	-
PreSep	<90

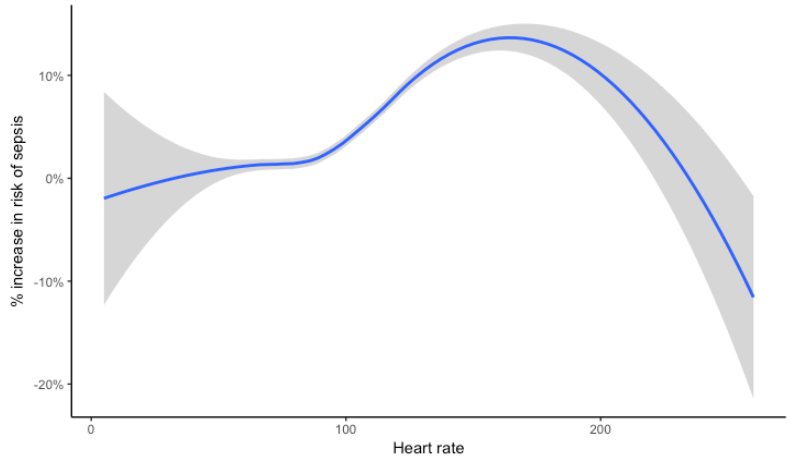
Table 7.18 Cut points for heart rate

Sepsis diagnosis		**lower cut point**	**upper cut point**
ED Doctor diagnosis	estimate	52	151
	AUC (95%CI)	*0.325 (0.295-0.354)*	*0.675 (0.646-0.705)*
Infection + 2SIRS	estimate	64	122
	AUC (95%CI)	*0.324 (0.31 - 0.338)*	*0.676 (0.662-0.69)*
SSC SOFA	estimate	52	150
	AUC (95%CI)	*0.418 (0.387-0.45)*	*0.582 (0.55-0.613)*
NICE (low, mod, high risk)	estimate	70	112
	AUC (95%CI)	*0.413 (0.401-0.424)*	*0.587 (0.576-0.599)*
NICE (mod & high risk)	estimate	65	120
	AUC (95%CI)	*0.328 (0.315-0.341)*	*0.672 (0.659-0.685)*
NICE (high risk only)	estimate	56	140
	AUC (95%CI)	*0.271 (0.25-0.239)*	*0.729 (0.707-0.75)*

ED – Emergency Department, SIRS – Systemic Inflammatory Response Criteria, SSC – Survive Sepsis Campaign, SOFA – Sequential Organ Failure Score, AUC – area under the curve

Figure 7.6 demonstrates that risk of sepsis is minimal when heart rate is below 100 beats per minute, but begins to increase at a rapid rate as heart rate increases beyond 100 beats per minute. Risk of sepsis peaks at around 160 beats per minute, and thereafter begins to decline. This decline in risk of sepsis at rates above 160 beats per minute may reflect that rates above 160 beats per minute are more likely to be associated with an electrophysiological origin than a compensatory response to low cardiac output.

Figure 7.6 Loess curve for heart rate (Pulse) vs NICE high sepsis risk



The reference range for heart rate is 60 to 100 beats per minute. Rates below 60 were combined into a single interval, while rates above 100 were established at 10 beat intervals up to 180 beats per minute. Rates above 180 were considered a single interval.

Systolic blood pressure

The normal physiological range for systolic blood pressure is 100 to 120mmHg.[197] **Table 7.19** identifies thresholds previously established in the literature, while **table 7.20** reports calculated cut points for different sepsis diagnoses with respect to systolic blood pressure.

Table 7.19 Existing thresholds for systolic blood pressure

	range
Normal physiological range	90 - 120
NICE Sepsis (low risk)	-
NICE Sepsis (mod risk)	91 - 100
NICE Sepsis (high risk)	<91
NEWS (0)	111-219
NEWS (1)	101 – 110
NEWS (2)	91-100
NEWS (3)	<91 or >219
UK Sepsis Trust (red flag)	<91
UK Sepsis Trust (amber flag)	91 - 100
Critical Illness Score (0)	>89
Critical Illness Score (1)	<90
Critical Illness Score (2)	-
PRESS (0)	>99
PRESS (1)	90 – 99
PRESS (2)	80 – 89
PRESS (3)	70 – 79
PRESS (4)	60 – 69
PRESS (5)	<60
PreSep	<90

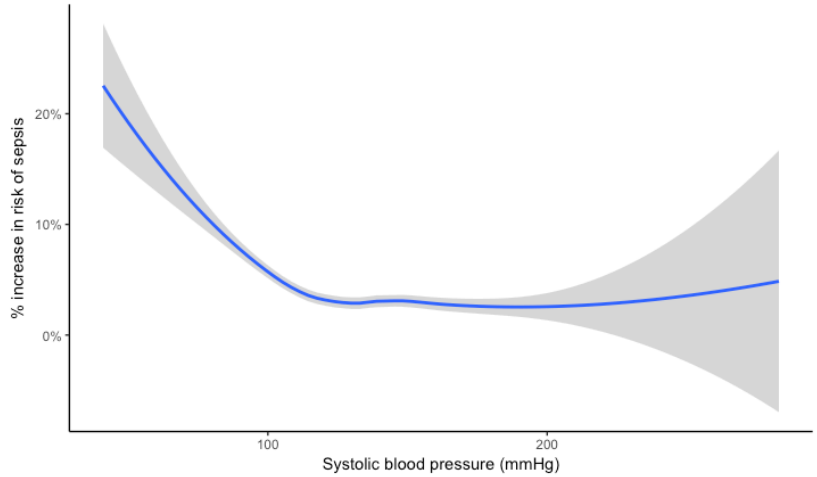
Table 7.20 Cut points for systolic blood pressure (SBP)

Sepsis diagnosis		**lower cut point**	**upper cut point**
ED Doctor diagnosis	estimate	82	194
	AUC (95%CI)	*0.599 (0.566-0.633)*	*0.401 (0.367-0.434)*
Infection + 2SIRS	estimate	100	169
	AUC (95%CI)	*0.537 (0.521-0.552)*	*0.463 (0.448-0.479)*
SSC SOFA	estimate	83	193
	AUC (95%CI)	*0.65 (0.618-0.681)*	*0.35 (0.319-0.382)*
NICE (low, mod, high risk)	estimate	109	157
	AUC (95%CI)	*0.55 (0.538-0.561)*	*0.45 (0.439-0.462)*
NICE (mod & high risk)	estimate	102	166
	AUC (95%CI)	*0.560 (0.546-0.574)*	*0.44 (0.426-0.454)*
NICE (high risk only)	estimate	88	186
	AUC (95%CI)	*0.601 (0.576-0.627)*	*0.399 (0.373-0.424)*

ED – Emergency Department, SIRS – Systemic Inflammatory Response Criteria, SSC – Survive Sepsis Campaign, SOFA – Sequential Organ Failure Score, AUC – area under the curve

Figure 7.7 depicts that risk of sepsis is minimal when systolic blood pressure is above 110mmHg (approximately), but begins to increase at a rapid rate as systolic pressure decreases below 100mmHg.

Figure 7.7 Loess curve for systolic blood pressure (SBP) vs NICE high sepsis risk



The reference range for systolic blood pressure is 100mmHg to 120mmHg. Below 100mmHg intervals were established every 10mmHg down to 60mmHg, and below 60mmHg were combined into a single interval. Above 120mmHg intervals were established every 10mmHg up to 180mmHg, above 180mmHg were combined into a single interval.

Glasgow Coma Score

The Glasgow Coma Score is a thirteen-point ordinal score, developed by Teasdale and Jennet, to assess consciousness in head injured patients.[199] A patient without altered consciousness would score 15 out of 15, consequently this is the normal physiological range.[199] **Table 7.21** identifies thresholds previously established in the literature, while **table 7.22** reports calculated cut points for different sepsis diagnoses with respect to Glasgow Coma Score.

Table 7.21 Existing thresholds for Consciousness (GCS)

	range
Normal physiological range	15
NICE Sepsis (low risk)	-
NICE Sepsis (mod risk)	-
NICE Sepsis (high risk)	<15
NEWS (0)	-
NEWS (1)	-
NEWS (2)	-
NEWS (3)	-
UK Sepsis Trust (red flag)	-
UK Sepsis Trust (amber flag)	-
Critical Illness Score (0)	15
Critical Illness Score (1)	8 – 14
Critical Illness Score (2)	<8
PRESS (0)	-
PRESS (1)	-
PRESS (2)	-
PRESS (3)	-
PRESS (4)	-
PRESS (5)	-
PreSep	<15

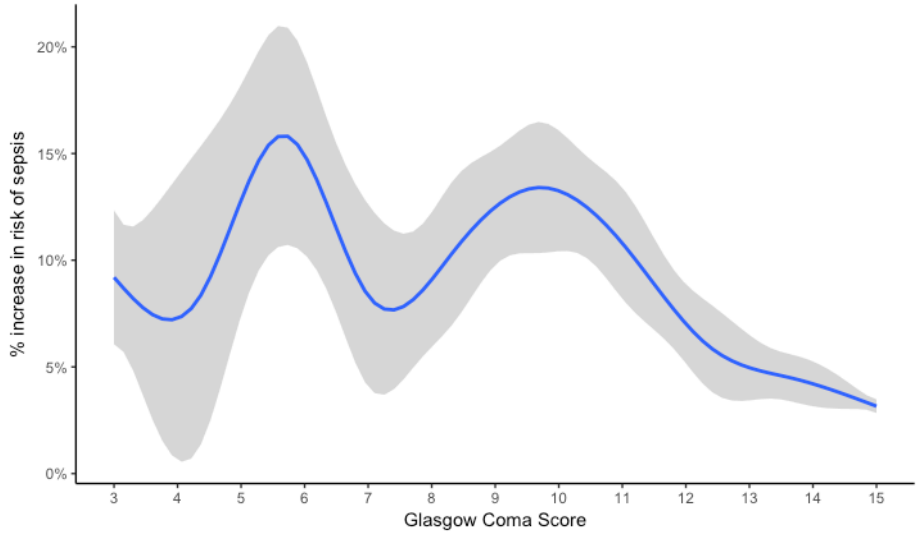
Table 7.22 Cut points for GCS (sum)

Sepsis diagnosis		lower cut point	upper cut point
ED Doctor diagnosis	estimate	8	15
	AUC (95%CI)	*0.601 (0.575-0.627)*	*0.399 (0.373-0.425)*
Infection + 2SIRS	estimate	13	15
	AUC (95%CI)	*0.548 (0.537-0.559)*	*0.452 (0.441-0.463)*
SSC SOFA	estimate	8	15
	AUC (95%CI)	*0.591 (0.566-0.616)*	*0.409 (0.384-0.434)*
NICE (low, mod, high risk)	estimate	14	15
	AUC (95%CI)	*0.543 (0.535-0.551)*	*0.457 (0.449-0.465)*
NICE (mod & high risk)	estimate	14	15
	AUC (95%CI)	*0.544 (0.535-0.554)*	*0.456 (0.446-0.465)*
NICE (high risk only)	estimate	10	15
	AUC (95%CI)	*0.570 (0.551-0.589)*	*0.430 (0.411-0.449)*

ED – Emergency Department, SIRS – Systemic Inflammatory Response Criteria, SSC – Survive Sepsis Campaign, SOFA – Sequential Organ Failure Score, AUC – area under the curve

Figure 7.8 illustrates that risk of sepsis fluctuates considerably, with peaks easily identified at approximately GCS10 and GCS6.

Figure 7.8 Loess curve for Glasgow Coma Score vs NICE high sepsis risk



The reference score for GCS is 15. As GCS reduces, the first peak on the Loess curve appears in the range 7-12. The second peak occurs at GCS below 7.

Ambulance staff commonly associate GCS below 9 with "unconsciousness". In addition, GCS 8 or below was calculated as a cut point, and has been utilised as a threshold in the Critical Illness Score. Consequently, intervals were established at GCS 15, 13-14, 9-12 and 8 or lower.

Temperature

The normal physiological range for body temperature is 36.6°C to 37.4°C.[197] **Table 7.22** identifies thresholds previously established in the literature, while **table 7.23** reports calculated cut points for different sepsis diagnoses with respect to temperature.

Table 7.23 Existing thresholds for temperature

	Range (°C)
Normal physiological range	36.5 – 38
NICE Sepsis (low risk)	-
NICE Sepsis (mod risk)	<36
NICE Sepsis (high risk)	-
NEWS (0)	36.1 - 38
NEWS (1)	35.1-36 or 38.1-39
NEWS (2)	>39
NEWS (3)	<35.1
UK Sepsis Trust (red flag)	-
UK Sepsis Trust (amber flag)	<36
Critical Illness Score (0)	-
Critical Illness Score (1)	-
Critical Illness Score (2)	-
PRESS (0)	Normal
PRESS (1)	-
PRESS (2)	-
PRESS (3)	hot
PRESS (4)	-
PRESS (5)	-
PreSep	<36 or >38

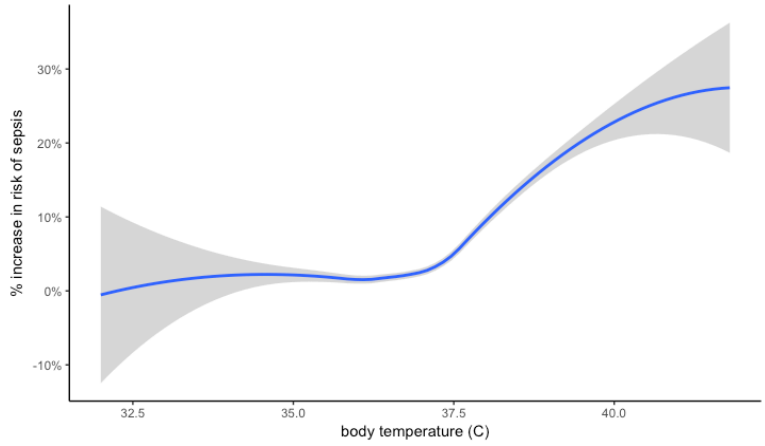
Table 7.24 Cut points for temperature

Sepsis diagnosis		**lower cut point** (°C)	**upper cut point** (°C)
ED Doctor diagnosis	estimate	34.9	39.2
	AUC (95%CI)	*0.232 (0.198-0.267)*	*0.768 (0.733-0.802)*
Infection + 2SIRS	estimate	35.8	37.8
	AUC (95%CI)	*0.302 (0.286-0.319)*	*0.698 (0.681-0.714)*
SSC SOFA	estimate	35	39.1
	AUC (95%CI)	*0.349 (0.313-0.385)*	*0.651 (0.615-0.687)*
NICE (low, mod, high risk)	estimate	36	37.4
	AUC (95%CI)	*0.326 (0.313-0.339)*	*0.674 (0.661-0.687)*
NICE (mod & high risk)	estimate	35.8	37.7
	AUC (95%CI)	*0.318 (0.302-0.334)*	*0.682 (0.666-0.698)*
NICE (high risk only)	estimate	35.2	38.7
	AUC (95%CI)	*0.288 (0.262-0.314)*	*0.712 (0.686-0.738)*

ED – Emergency Department, SIRS – Systemic Inflammatory Response Criteria, SSC – Survive Sepsis Campaign, SOFA – Sequential Organ Failure Score, AUC – area under the curve

Figure 7.9 depicts that risk of sepsis is minimal when body temperature is below 37.5°C (approximately), but begins to increase at a rapid rate as body temperature increases above 37.5°C.

Figure 7.9 Loess curve for body temperature vs NICE high sepsis risk



The reference range for body temperature is 36.6°C to 37.4C. Below 36.6°C intervals were established every 0.5°C down to 35.0°C, and below 35.0°C were

combined into a single interval. Above 37.4°C, intervals were established every 0.5°C up to 40.0°C, and above 40.0°C were combined into a single interval.

7.3.3.1 Regression of multivariate models to generate scores

Three multivariable models, utilising different measures of consciousness, were evaluated. **Tables 7.25, 7.26** and **7.27** report the regression coefficients, 95% confidence interval and p value for each variable used, for each of the three models. Scores are assigned by rounding statistically significant regression coefficients to the nearest integer. For example, in **table 7.25** the regression coefficient for Age "60 to 69" is 0.82. This is rounded up to 1. Similarly, in **table 7.26** the regression coefficient for "pallid" skin is 0.40. This is rounded down to a score of 0 despite being highly significant (p<0.001). Non-significant regression coefficients are highlighted, and all are allocated a score of 0.

Table 7.25 Regression of Model A (utilising GCS sum)

		$ß_i$	lo 95	hi 95	p-value	Score assigned
	intercept	-5.46	-5.98	-4.93	0.00	
Location	home	reference				0
	nursing home	0.21	-0.14	0.56	0.23	0
	other	-0.28	-0.64	0.08	0.13	0
Age	below 40	reference				0
	40 to 49	0.12	-0.58	0.81	0.74	0
	50 to 59	0.40	-0.20	1.00	0.19	0
	60 to 69	0.82	0.30	1.33	<0.001	1
	70 to 79	0.94	0.45	1.43	<0.001	1
	80 to 89	0.87	0.37	1.36	<0.001	1
	90 to 99	1.35	0.79	1.92	<0.001	1
	100 plus	1.59	-0.17	3.35	0.08	0
Respirations	below 10	-13.20	-948.66	922.25	0.98	0
	10 to 20	reference				0
	21 to 25	0.52	0.19	0.86	<0.001	1
	26 to 30	0.89	0.56	1.22	<0.001	1
	31 to 35	1.00	0.57	1.44	<0.001	1
	36 to 40	1.28	0.91	1.65	<0.001	1
	41 to 45	2.14	1.46	2.81	<0.001	2
	46 to 50	0.87	-0.03	1.76	0.06	0
	51 to 55	1.27	-0.17	2.70	0.08	0
	56 to 60	1.96	1.02	2.91	<0.001	2
	60 plus	-12.60	-3227.5	3202.2	0.99	0
SpO2	above 93	reference				0

	89 to 93	0.92	0.64	1.20	<0.001	1
	85 to 88	0.79	0.39	1.19	<0.001	1
	below 85	1.19	0.84	1.55	<0.001	1
Pulse	below 60	-0.58	-1.43	0.27	0.18	0
	60 to 100		reference			0
	101 to 110	0.62	0.29	0.94	<0.001	1
	111 to 120	0.63	0.29	0.97	<0.001	1
	121 to 130	0.96	0.58	1.34	<0.001	1
	131 to 140	0.88	0.42	1.33	<0.001	1
	141 to 150	1.84	1.34	2.35	<0.001	2
	151 to 160	1.11	0.39	1.83	<0.001	1
	161 to 170	0.38	-0.86	1.63	0.55	0
	171 to 180	0.04	-2.11	2.18	0.97	0
	180 plus	1.07	-0.13	2.28	0.08	0
SBP	below 60	0.74	-1.35	2.82	0.49	0
	60 to 69	1.39	0.43	2.36	<0.001	1
	70 to 79	1.04	0.26	1.82	0.01	1
	80 to 89	0.72	0.19	1.25	0.01	1
	90 to 99	0.65	0.24	1.05	<0.001	1
	100 to 120		reference			0
	121 to 129	-0.06	-0.44	0.32	0.74	0
	130 to 139	-0.09	-0.46	0.28	0.64	0
	140 to 149	-0.21	-0.60	0.18	0.30	0
	150 to 159	-0.05	-0.46	0.36	0.80	0
	160 plus	-0.72	-1.13	-0.32	<0.001	-1
GCS (sum)	15		reference			0
	13 to 14	-0.25	-0.62	0.12	0.19	0
	9 to 12	0.67	0.27	1.08	<0.001	1
	3 to 8	0.48	-0.08	1.05	0.09	0
Temperature	below 35.0	-0.49	-1.49	0.50	0.33	0
	35.0 to 35.5	0.18	-0.41	0.76	0.55	0
	35.6 to 36.0	-0.24	-0.70	0.22	0.31	0
	36.1 to 36.5	-0.21	-0.56	0.15	0.25	0
	36.6 to 37.4		reference			0
	37.5 to 38.0	0.55	0.20	0.91	<0.001	1
	38.1 to 38.5	1.16	0.78	1.54	<0.001	1
	38.6 to 39.0	1.23	0.79	1.66	<0.001	1
	39.1 to 39.5	1.09	0.53	1.64	<0.001	1
	39.6 to 40.0	1.65	1.07	2.23	<0.001	2
	above 40.0	1.91	1.09	2.74	<0.001	2
Skin	normal		reference			0
	cyanosed	-0.05	-0.68	0.58	0.88	0
	flushed	0.13	-0.25	0.50	0.50	0
	jaundice	0.90	0.01	1.79	0.05	1

	mottling	0.82	-0.33	1.96	0.16	0
	pallor	0.40	0.14	0.66	<0.001	0
	rash	-12.76	-1121.5	1095.9	0.98	0
CBRT	normal	reference				0
	delayed	-0.06	-0.42	0.31	0.75	0

Highlighted estimates indicate non-significant estimates

Table 7.26 Regression of Model B (utilising GCS components)

		ßi	**lo 95**	**hi 95**	**p-value**	**Score assigned**
	Intercept	-5.46	-5.98	-4.93	0.00	
Location	home	reference				0
	nursing home	0.20	-0.14	0.55	0.25	0
	other	-0.28	-0.64	0.07	0.12	0
Age	under 40	reference				0
	40 to 49	0.10	-0.60	0.79	0.79	0
	50 to 59	0.39	-0.21	0.99	0.20	0
	60 to 69	0.79	0.28	1.31	<0.001	1
	70 to 79	0.93	0.44	1.42	<0.001	1
	80 to 89	0.84	0.34	1.33	<0.001	1
	90 to 99	1.33	0.77	1.90	<0.001	1
	100 plus	1.63	-0.07	3.34	0.06	0
Respirations	below 10	-13.47	-940.42	913.48	0.98	0
	10 to 20	reference				0
	21 to 25	0.53	0.20	0.87	<0.001	1
	26 to 30	0.89	0.56	1.22	<0.001	1
	31 to 35	1.01	0.57	1.45	<0.001	1
	36 to 40	1.30	0.92	1.67	<0.001	1
	41 to 45	2.12	1.45	2.80	<0.001	2
	46 to 50	0.85	-0.06	1.75	0.07	0
	51 to 55	1.21	-0.22	2.64	0.10	0
	56 to 60	1.94	1.00	2.88	<0.001	2
	60 plus	-12.59	-3230.28	3205.09	0.99	0
SpO2	above 93	reference				0
	89 to 93	0.91	0.63	1.19	<0.001	1
	85 to 88	0.79	0.38	1.19	<0.001	1
	below 85	1.18	0.83	1.53	<0.001	1
Pulse	below 60	-0.58	-1.43	0.27	0.18	0
	60 to 100	reference				0
	101 to 110	0.64	0.31	0.96	<0.001	1
	111 to 120	0.63	0.29	0.97	<0.001	1
	21 to 130	0.96	0.59	1.34	<0.001	1
	131 to 140	0.87	0.42	1.32	<0.001	1
	141 to 150	1.84	1.33	2.34	<0.001	2

	151 to 160	1.09	0.37	1.81	<0.001	1
	161 to 170	0.45	-0.78	1.67	0.47	0
	171 to 180	0.01	-2.13	2.16	0.99	0
	180 plus	1.05	-0.16	2.27	0.09	0
SBP	below 60	0.87	-1.11	2.85	0.39	0
	60 to 69	1.44	0.48	2.40	<0.001	1
	70 to 79	1.06	0.29	1.83	0.01	1
	80 to 89	0.71	0.18	1.25	0.01	1
	90 to 99	0.64	0.24	1.05	<0.001	1
	100 to 120		reference			0
	121 to 129	-0.07	-0.45	0.31	0.71	0
	130 to 139	-0.08	-0.46	0.29	0.66	0
	140 to 149	-0.19	-0.59	0.20	0.33	0
	150 to 159	-0.07	-0.48	0.34	0.75	0
	160 plus	-0.72	-1.13	-0.32	<0.001	-1
GCS (verbal)	oriented		reference			0
	confused	0.02	-0.33	0.36	0.92	0
	Inapp. words	0.50	-0.25	1.25	0.19	0
	Incomp. sounds	0.46	-0.21	1.13	0.18	0
	no response	0.83	0.31	1.34	<0.001	1
Temperature	below 35	-0.50	-1.49	0.48	0.32	0
	35.0 to 35.5	0.17	-0.42	0.75	0.58	0
	35.6 to 36.0	-0.25	-0.71	0.21	0.28	0
	36.1 to 36.5	-0.21	-0.56	0.15	0.25	0
	36.6 to 37.4		reference			0
	37.5 to 38.0	0.55	0.19	0.90	<0.001	1
	38.1 to 38.5	1.15	0.78	1.53	<0.001	1
	38.6 to 39.0	1.22	0.78	1.66	<0.001	1
	39.1 to 39.5	1.09	0.53	1.64	<0.001	1
	39.6 to 40.0	1.62	1.04	2.20	<0.001	2
	above 40	1.89	1.06	2.72	<0.001	2
Skin	normal		reference			0
	cyanosed	0.00	-0.63	0.62	1.00	0
	flushed	0.13	-0.24	0.51	0.48	0
	jaundice	0.89	0.00	1.78	0.05	1
	mottling	0.79	-0.36	1.93	0.18	0
	pallor	0.40	0.14	0.66	<0.001	0
	rash	-12.76	-1121.39	1095.86	0.98	0
CBRT	normal		reference			0
	delayed	-0.06	-0.42	0.31	0.76	0

Highlighted estimates indicate non-significant estimates

Table 7.27 Regression of Model C (utilising AVPU)

		$ß_i$	lo 95	hi 95	p-value	Score assigned
	intercept	-5.50	-6.02	-4.97	0.00	
Location	home		reference			0
	nursing home	0.24	-0.09	0.58	0.15	0
	other	-0.28	-0.64	0.07	0.12	0
Age	below 40		reference			0
	40 to 49	0.14	-0.55	0.84	0.69	0
	50 to 59	0.45	-0.16	1.05	0.15	0
	60 to 69	0.83	0.32	1.34	<0.001	1
	70 to 79	0.98	0.49	1.47	<0.001	1
	80 to 89	0.88	0.38	1.37	<0.001	1
	90 to 99	1.37	0.81	1.93	<0.001	1
	100 plus	1.34	-0.40	3.08	0.13	0
Respirations	below 10	-13.25	-945.34	918.84	0.98	0
	10 to 20		reference			0
	21 to 25	0.54	0.21	0.87	<0.001	1
	26 to 30	0.91	0.58	1.24	<0.001	1
	31 to 35	1.01	0.57	1.45	<0.001	1
	36 to 40	1.32	0.95	1.69	<0.001	1
	41 to 45	2.15	1.49	2.80	<0.001	2
	46 to 50	0.80	-0.10	1.69	0.08	0
	51 to 55	1.23	-0.18	2.64	0.09	0
	56 to 60	1.99	1.05	2.93	<0.001	2
	60 plus	-12.55	-3229.45	3204.36	0.99	0
SpO2	above 93		reference			0
	89 to 93	0.90	0.62	1.18	<0.001	1
	85 to 88	0.76	0.35	1.16	<0.001	1
	below 85	1.16	0.81	1.50	<0.001	1
Pulse	below 60	-0.58	-1.42	0.27	0.18	0
	60 to 100		reference			0
	101 to 110	0.62	0.30	0.95	<0.001	1
	111 to 120	0.64	0.30	0.98	<0.001	1
	121 to 130	0.99	0.61	1.36	<0.001	1
	131 to 140	0.89	0.45	1.33	<0.001	1
	141 to 150	1.88	1.38	2.37	<0.001	2
	151 to 160	1.14	0.43	1.85	<0.001	1
	161 to 170	0.74	-0.39	1.87	0.20	0
	171 to 180	0.07	-2.07	2.21	0.95	0
	180 plus	1.12	-0.08	2.33	0.07	0
SBP	below 60	0.72	-1.33	2.78	0.49	0
	60 to 69	1.40	0.45	2.36	<0.001	1
	70 to 79	1.01	0.23	1.79	0.01	1
	80 to 89	0.73	0.20	1.26	0.01	1

	90 to 99	0.64	0.24	1.04	<0.001	1
	100 to 120		reference			0
	121 to 129	-0.10	-0.47	0.28	0.62	0
	130 to 139	-0.11	-0.48	0.27	0.58	0
	140 to 149	-0.20	-0.59	0.19	0.30	0
	150 to 159	-0.08	-0.49	0.32	0.69	0
	160 plus	-0.71	-1.11	-0.31	<0.001	-1
AVPU	alert		reference			0
	verbal	0.44	0.08	0.80	0.02	0
	pain	0.48	-0.12	1.07	0.12	0
	no response	0.61	-0.13	1.36	0.11	0
Temperature	below 35.0	-0.50	-1.49	0.49	0.32	0
	35.0 to 35.5	0.13	-0.45	0.71	0.66	0
	35.6 to 36.0	-0.28	-0.74	0.18	0.23	0
	36.1 to 36.5	-0.19	-0.54	0.16	0.28	0
	36.6 to 37.4		reference			0
	37.5 to 38.0	0.53	0.17	0.88	<0.001	1
	38.1 to 38.5	1.17	0.79	1.54	<0.001	1
	38.6 to 39.0	1.17	0.74	1.60	<0.001	1
	39.1 to 39.5	1.13	0.59	1.67	<0.001	1
	39.6 to 40.0	1.60	1.02	2.18	<0.001	2
	above 40.0	1.90	1.07	2.73	<0.001	2
Skin	normal		reference			0
	cyanosed	-0.10	-0.73	0.53	0.76	0
	flushed	0.12	-0.26	0.49	0.54	0
	jaundice	0.82	-0.08	1.71	0.07	0
	mottling	0.99	-0.06	2.05	0.07	0
	pallor	0.39	0.13	0.65	<0.001	0
	rash	-12.73	-1122.95	1097.49	0.98	0
CBRT	normal		reference			0
	delayed	-0.04	-0.40	0.32	0.82	0

Highlighted estimates indicate non-significant estimates

7.3.3.2 Pragmatic adjustments of models

The intervals chosen when subdividing each continuous variable will ultimately dictate the calculated regression coefficients ($ß_i$). The calculated regression coefficients ($ß_i$) are utilised to generate scores. It stands to reason therefore, that the intervals chosen will significantly impact the scores generated, and ultimately ease of use of the final screening tool. To minimise loss of precision, small intervals have been employed in the preceding multivariable regressions. This has resulted in the generation of scores, that although more precise, would also be more complex for the end-user. For example, if we were to consider the

regression coefficients ($ß_i$) generated in **table 7.25,** we assign scores for respiratory rate as reported in **table 7.28**.

Table 7.28 Scores for respirations in Model A

Variable	Respiratory rate	Score
Resps	20 or lower	0
	21 to 40	1
	41 to 45	2
	46 to 55	0
	56 to 60	2
	61 or over	0

Examination of **figure 7.3** suggests that risk of sepsis increases at a steady rate when Respirations are in the range 20 to 60, consequently we would not expect respiratory rates in the interval "46 to 50" to generate a score of "0". Inspection of **table 7.25** reveals that, in Model A, for the variable Resps, the interval "46 to 50" has a regression coefficient ($ß_i$) of 0.87 and p value of 0.06, while the interval "51 to 55" has a regression coefficient ($ß_i$) of 1.27 and p value of 0.08. Both intervals approach, but fail to achieve significance, and are thus automatically assigned a score of "0". However, if the intervals for Resps for ModelA are redefined to merge the "41 to 45" and "46 to 50" into a single "41 to 50" interval, the resulting regression coefficient ($ß_i$) is 1.66 and p value is less than 0.001. Similarly, if the intervals for Resps are redefined to merge the "51 to 55" and "56 to 60" into a single "51 to 60" interval, the regression coefficient ($ß_i$) is 1.78 and p value is less than 0.001. The resulting adjusted scores for respirations are reported in table **7.29**.

Table 7.29 Adjusted scores for respirations

Variable	Respiratory rate	Score
Resps	20 or lower	0
	21 to 40	1
	41 to 60	2
	61 or 0ver	0

This simplification of the model was undertaken to reduce complexity, and to facilitate ease of use by ambulance clinicians. The simplification is justified through the sustained statistical significance; however, it is acknowledged that the simplification may lead to a small loss of precision in the model.

Four variables, utilised across all three models, have been subject to pragmatic adjustment to facilitate ease of use. These variables include **Age**, **Resps**, **Pulse** and **Skin**. Adjustments for **Age** and **Pulse** are outlined below in **tables 7.30**, **7.31**, **7.32** and **7.33**. Regression tables, upon which these adjustments are predicated, are reported in **Appendix 3**.

Table 7.30 Scores for Age before adjustment

Variable	Age	Score
Age	Under 60	0
	60 to 99	1
	100 or older	0

Table 7.31 Scores for Age following adjustment

Variable	Age	Score
Age	Under 60	0
	60 or older	1

Table 7.32 Scores for Pulse before adjustment

Variable	Pulse rate	Score
Pulse	100 or lower	0
	101 to 140	1
	141 to 150	2
	151 to 160	1
	161 or 0ver	0

Table 7.33 Scores for Pulse following adjustment

Variable	Pulse rate	Score
Pulse	100 or lower	0
	101 to 140	1
	141 to 160	2
	161 or 0ver	0

Regression models suggest that the only category of Skin that correlated with high risk of sepsis, and produced a regression coefficient ($ß_i$) sufficiently high to generate a score is "jaundice". These same regression models indicate that the category "pallor" correlated with sepsis in some models, but the regression coefficient ($ß_i$) was small and generated a score of zero. Furthermore, the category "mottling" approached, but failed to achieve statistical significance. As with the Age variable, failure to achieve statistical significance may be related to the relatively small numbers of cases in each category (see **table 7.34)**.

Table 7.34 Cases of high risk sepsis according to `Skin` category

	cyanosed	**flushed**	**jaundice**	**mottling**	**normal**	**pallor**	**rash**
no sepsis	158	798	83	44	10126	2484	21
sepsis	23	58	9	9	240	185	0

Previously published sepsis assessment tools have suggested that pale or mottled skin should be considered an indicator for risk of sepsis.[5] A pragmatic adjustment, predicted on a combination of clinical reasoning, and the low number of cases reported in potentially important skin findings, resulted in the `Skin` categories "jaundice", "pallor" and "mottling" being merged into a single category. The results of this adjustment are reported in **tables 7.35** and **7.36**.

Table 7.35 Categories for `Skin` before adjustment

Category	**ß$_i$**	**p-value**	**score**
normal		[reference]	
cyanosed	-0.05	0.88	0
flushed	0.13	0.50	0
jaundice	0.90	0.05	1
mottling	0.82	0.16	0
pallor	0.40	<0.001	0
rash	-12.76	0.98	0

Table 7.36 Categories for `Skin` after adjustment

Category	**ß$_i$**	**p-value**	**score**
normal		[reference]	
jaundice, pallor, mottling	0.51	<0.001	1
any other	0.23	0.14	0

GCS (sum) is only utilised in one of the three models (Model A). The intervals chosen for the original regression were informed by previous studies and accepted clinical norms. The interval "9 to 12" achieved statistical significance, and generated a score of 1, while the interval "3 to 8" approached, but failed to achieve statistical significance. Had the result been statistically significant, the regression coefficient (ß$_i$) of 0.48 comes close to being assigned a score of "1". As with `Age`, and `Skin`, the number of cases of high risk sepsis in important

categories is relatively low. Failure to achieve statistical significance may reflect too few cases (see **table 7.37**).

Table 7.37 Cases of high risk sepsis according to `GCS_sum` category

	3 to 8	**9 to 12**	**13 to 14**	**15**
no sepsis	330	561	1373	12992
sepsis	34	67	61	421

Previously published sepsis assessment tools have suggested that reduced GCS, including scores in the interval "3 to 8" should be considered an indicator for risk of sepsis.[107] If the categories for `GCS_sum` are redefined to merge the "3 to 8" and "9 to 12" into a single "3 to 12" category, the resulting regression coefficient ($ß_i$) is 0.78 and p value is less than 0.001. This suggests that it may be appropriate to merge these two categories. However, unlike the previously listed adjustments, the GCS_sum is the sum of three ordered categorical variables, and so is not a typical continuous variable, despite being reported as a numeric value. In other words, the "difference" between GCS scores of 4 and 5 is not necessarily equivalent to the difference between scores of 14 and 15. As a consequence, it is less clear whether it is clinically or statistically appropriate to merge the categories further ("3 to 12", "13 to 15") or to remain as above ("3 to 8", "9 to 12", "13 to 15"). Consequently, an additional model (Model A2) with the merged categories has been specified.

7.3.3.3 Model simplification

Each of the adjusted models can be simplified by merging together those intervals, within each variable, that have a consistent score. The scores associated with each model are reported below in **tables 7.38**. **7.39**, **7.40** and **7.41**.

Table 7.38 Final scoring Model A (GCS 9-12 only)

Variable	Interval	Score
Age	Under 60	0
	60 or older	1
Respirations	20 or lower	0
	21 to 40	1
	41 o 60	2
	61 or over	0
SpO2	94 or higher	0
	93 or lower	1
Pulse	100 or lower	0
	101 to 140	1
	141 to 160	2
	161 or over	0
SBP	Under 60	0
	60 to 99	1
	100 to 159	0
	160 or higher	-1
Temperature	37.4 or lower	0
	37.5 to 39.5	1
	39.6 or higher	2
Skin	Jaundice, pallor, mottling	1
	Any other	0
GCS (sum)	13 to 15	0
	9 to 12	1
	8 or lower	0

Table 7.39 Final scoring Model A2 (GCS<=12)

Variable	Interval	Score
Age	Under 60	0
	60 or older	1
Respirations	20 or lower	0
	21 to 40	1
	41 to 60	2
	61 or over	0
SpO2	94 or higher	0
	93 or lower	1
Pulse	100 or lower	0
	101 to 140	1
	141 to 160	2
	161 or over	0
SBP	Under 60	0
	60 to 99	1
	100 to 159	0
	160 or higher	-1
Temperature	37.4 or lower	0
	37.5 to 39.5	1
	39.6 or higher	2
Skin	Jaundice, pallor, mottling	1
	Any other	0
GCS (sum)	13 to 15	0
	3 to 12	1

Table 7.40 Final scoring Model B (GCS verbal)

Variable	Interval	Score
Age	Under 60	0
	60 or older	1
Respirations	20 or lower	0
	21 to 40	1
	41 to 60	2
	61 or over	0
SpO2	94 or higher	0
	93 or lower	1
Pulse	100 or lower	0
	101 to 140	1
	141 to 160	2
	161 or over	0
SBP	Under 60	0
	60 to 99	1
	100 to 159	0
	160 or higher	-1
Temperature	37.4 or lower	0
	37.5 to 39.5	1
	39.6 or higher	2
Skin	Jaundice, pallor, mottling	1
	Any other	0
GCS (verbal)	Any response	0
	No response	1

Table 7.41 Final scoring Model C (AVPU)

Variable	Interval	Score
Age	Under 60	0
	60 or older	1
Respirations	20 or lower	0
	21 to 40	1
	41 to 60	2
	61 or over	0
SpO2	94 or higher	0
	93 or lower	1
Pulse	100 or lower	0
	101 to 140	1
	141 to 160	2
	161 or over	0
SBP	Under 60	0
	60 to 99	1
	100 to 159	0
	160 or higher	-1
Temperature	37.4 or lower	0
	37.5 to 39.5	1
	39.6 or higher	2

7.3.3.4 Model comparison

Four different models, to identify high risk sepsis, have been proposed. To determine which model to carry forward, each model is compared utilising the AIC statistic, to determine which model provides the best fit to the derivation dataset. The AIC statistic for each model is reported in **table 7.42**.

Table 7.42 Comparison of models

Model	AIC statistic
Model A	2856.1
Model A2	2854.1
Model B	2864.4
Model C	3315.0

AIC statistics for Model A, Model A2 and Model B are similar, while that of Model C is a little higher. Technically, Model A2 has the lowest AIC statistic suggesting it is the best fitting model. However, in clinical practice there may not be a

noticeable difference in performance between Model A, Model A2 and Model B. To determine which of the four different models best predicts the likelihood of high sepsis risk we can consider the Brier Score for each model.

Brier Scores were originally developed to assess the accuracy of weather forecasts. However, they are commonly used in the assessment of accuracy of probabilistic forecasts. The Brier score is calculated utilising the squared error of a probabilistic forecast. Put simply, the Brier Score is a measure of how well your prediction model compares to the true outcome. A Brier Score of "0" indicates perfect prediction, while a score of "1" indicates a wholly inaccurate prediction. The Brier score was used to quantify how well each model predicted the outcome of high risk of sepsis.

Table 7.43 Comparison of models

Model	Brier Score
Model A	0.03212072
Model A2	0.03207829
Model B	0.03209519
Model C	0.03248747

Brier Scores reported in **table 7.43** indicate there is little difference between model performance with respect to accuracy. Model A2, utilising `GCS_sum` as the measure of consciousness, with intervals established between "3 to 12", "13 to 14" and "15", is technically the most appropriate model to take forward. Of the four models proposed, Models B and C were excluded due to their high AIC scores. Model A2 was a simplification of Model A, without any observed loss of performance. Model A2 also has both the lowest AIC statistic and Brier Score, indicating that it is the model with the best fit for the data and is the most accurate.

7.3.4 Model estimation

The final regression model (Model A2) used to generate the SEPSIS score, including regression coefficients for the merged intervals is fully reported in **table 7.44.**

Table 7.44 Regression model following pragmatic adjustments

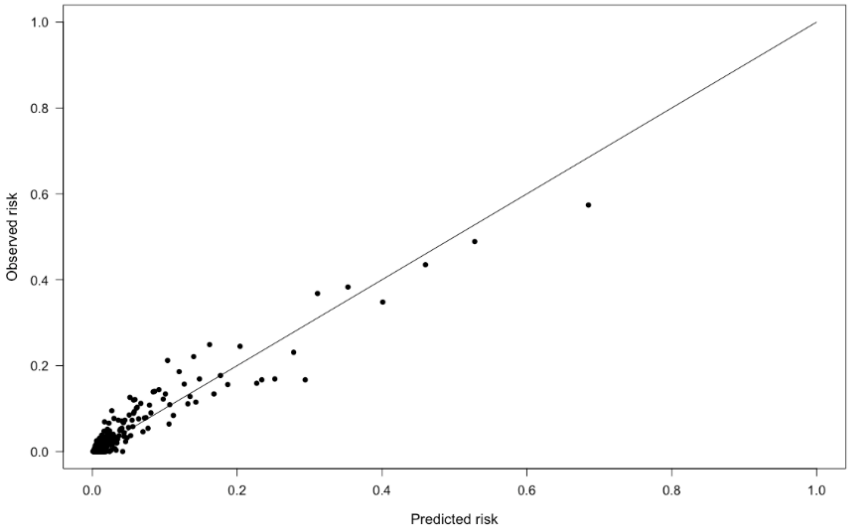
		ßi	lo 95	hi 95	p-value	Score assigned
	intercept	-5.46	-5.98	-4.93	0.00	
Age	below 40	reference				0
	40 to 60	0.35	-0.15	0.86	0.17	0
	over 60	0.94	0.51	1.36	<0.001	1
Respirations	Below 10	-12.44	-559.61	534.73	0.96	0
	10 to 20	reference				0
	21 to 40	0.90	0.66	1.14	<0.001	1
	40 to 60	1.72	1.26	2.18	<0.001	2
	60 plus	-11.77	-1970.04	1946.51	0.99	0
SpO2	above 93	reference				0
	below 94	1.03	0.80	1.26	<0.001	1
Pulse	below 60	-0.56	-1.40	0.28	0.19	0
	60 to 100	reference				0
	101 to 140	0.75	0.51	0.99	<0.001	1
	141 to 160	1.67	1.27	2.08	<0.001	2
	over 160	0.60	-0.16	1.35	0.12	0
SBP	below 60	0.50	-1.53	2.52	0.63	0
	60 to 99	0.65	0.33	0.97	<0.001	1
	100 to 120	reference				0
	121 to 160	-0.21	-0.47	0.05	0.11	0
	over 160	-0.72	-1.10	-0.34	<0.001	-1
GCS (sum)	15	reference				0
	13 to 15	-0.13	-0.48	0.21	0.45	0
	3 to 12	0.78	0.47	1.09	<0.001	1
Temperature	below 36.6	-0.20	-0.48	0.09	0.18	0
	36.6 to 37.4	reference				0
	37.5 to 39.5	0.97	0.71	1.23	<0.001	1
	above 39.5	1.71	1.23	2.18	<0.001	2
Skin	normal	reference				0
	jaundice, pallor, mottling	0.51	0.27	0.75	<0.001	1
	any other	0.23	-0.08	0.55	0.14	0
				Maximum score		11

7.3.5 Model performance

7.3.5.1 Model calibration

When calculated on the derivation sample, the Hosmer-Lemeshow goodness-of-fit test demonstrated statistical evidence of inadequate fit (λ^2= 1443.1; p<0.001). However, as previously identified, a significant Hosmer-Lemeshow statistic is common when using large datasets. The alternate means of assessing overfitting, the calibration slope (1.0), suggested little overfitting (see **figure 7.10**).

Figure 7.10 Calibration plot for Model A2



Further evidence of adequate model specification is seen by examining the observed versus predicted outcomes (based on predicted risk) generated when calculating the Hosmer-Lemeshow statistic (see **table 7.45**).

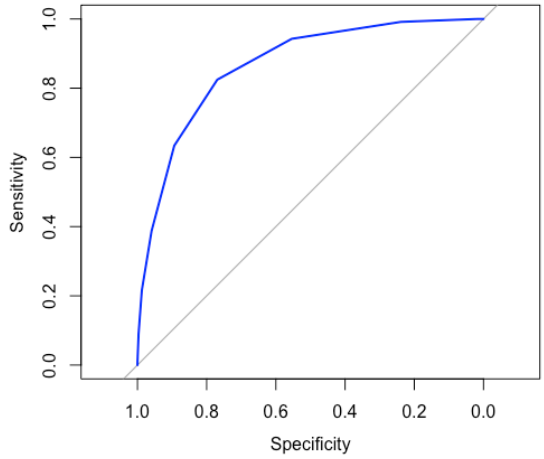
Table 7.45 Hosmer-Lemeshow observed vs predicted risk

Estimated risk	Observed		Predicted	
Interval	no sepsis	sepsis	no sepsis	sepsis
1.55e-08 - 0.00394	1436	2	1434	4
0.00394 - 0.00531	1466	1	1460	7
0.00531 - 0.00692	1497	8	1495	10
0.00692 - 0.00874	1162	5	1157	10
0.00874 - 0.0112	1171	7	1166	12
0.0112 - 0.017	1227	11	1221	17
0.017 - 0.0238	1238	31	1243	26
0.0238 - 0.0423	1275	48	1281	42
0.0423 - 0.0987	1204	112	1232	84
0.0987 - 0.879	1000	289	987	302

7.3.5.2 Model discrimination

Area under the receiver operating characteristic curve was 0.87 (95% CI 0.85 to 0.88) for the outcome of high sepsis risk at hospital (see **figure 7.11**).

Figure 7.11 Area under the Curve



7.3.5.3 Model operating characteristics

Several measures of model performance were calculated including sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio and diagnostic odds ratio for each point score. Performance characteristics of the SEPSIS score among adult medical patients is reported in **table 7.46**, while performance in patients with infection is reported in **table 7.47**. **Figure 7.12** graphically depicts probability of sepsis for each point on the SEPSIS score.

Figure 7.12 Observed vs expected probability of sepsis

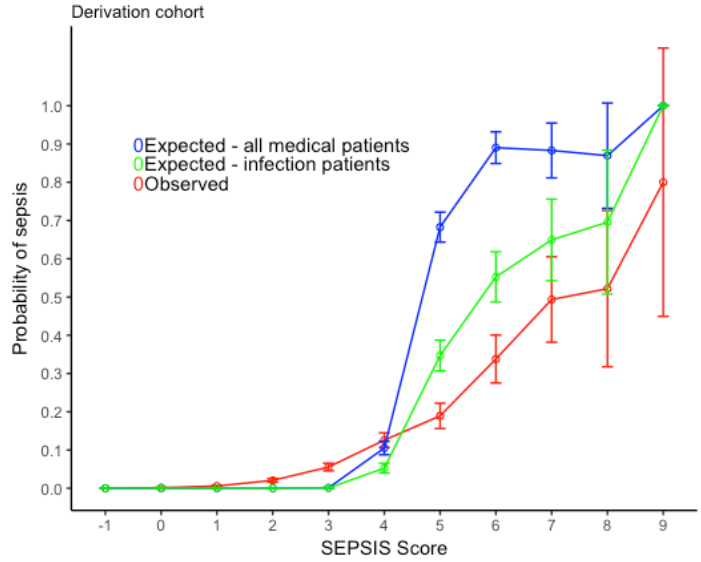


Table 7.46 Operating characteristics for the SEPSIS score among patients with undifferentiated medical complaints

score	≥1	≥2	≥3	≥4	≥5	≥6	≥7	≥8	≥9
Outcome									
TP	588	559	489	376	230	128	54	16	4
FP	111776	6898	3570	1647	633	196	51	12	1
TN	3694	8572	11900	13823	14837	15274	15419	15458	15469
FN	5	34	104	217	363	465	539	577	593
Operating characteristics									
Sens	0.99 (0.98-1.00)	0.94 (0.92-0.96)	0.82 (0.79-0.85)	0.63 (0.59-0.67)	0.39 (0.35-0.43)	0.22 (0.18-0.25)	0.09 (0.11-0.17)	0.03(0.02-0.04)	0.01 (0.0-0.02)
Spec	0.24 (0.23-0.25)	0.55 (0.55-0.56)	0.77 (0.76-0.78)	0.89 (0.89-0.90)	0.96 (0.96-0.96)	0.99 (0.99-0.99)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)
PPV	0.05 (0.04-0.05)	0.07 (0.07-0.08)	0.12 (0.11-0.13)	0.19 (0.17-0.20)	0.27 (0.24-0.30)	0.40 (0.34-0.45)	0.51 (0.41-0.61)	0.57 (0.37-0.76)	0.80 (0.28-0.99)
NPV	1.00(1.00-1.00)	1.00(0.99-1.00)	0.99 (0.99-1.00)	0.98 (0.98-0.99)	0.98 (0.97-0.98)	0.97 (0.97-0.97)	0.97 (0.96-0.97)	0.96 (0.96-0.97)	0.96 (0.96-0.97)
PLR	1.30 (1.29-1.32)	2.11 (2.06-2.17)	3.57 (3.41-3.75)	5.96 (5.52-6.43)	9.48 (8.35-10.76)	17.0 (13.9-21.0)	27.6 (19.01-40.1)	34.8 (16.5-73.2)	104 (11.7-932.2)
NLR	0.04 (0.01-0.08)	0.10 (0.07-0.14)	0.23 (0.19-0.27)	0.41 (0.37-0.46)	0.64 (0.60-0.68)	0.79 (0.76-0.83)	0.91 (0.89-0.94)	0.97 (0.96-0.99)	0.99 (0.98-1.0)
DOR	36.9 (15.3-89.0)	20.4 (14.4-28.9)	15.7 (12.6-19.4)	14.5 (12.2-17.3)	14.9 (12.4-17.8)	21.5 (16.9-27.3)	30.3 (20.5-44.8)	35.7 (16.9-75.6)	105 (11.7-941.3)

Sens-sensitivity, Spec-specificity, PPV-positive predictive value, NPV-negative predictive value, PLR-positive likelihood ratio, NLR-negative likelihood ratio, DOR-diagnostic odds ratio
Items in brackets () 95% confidence interval

Table 7.47 Operating characteristics for the SEPSIS score among patients with undifferentiated medical complaints

score	≥1	≥2	≥3	≥4	≥5	≥6	≥7	≥8	≥9
Outcome									
TP	588	559	489	376	230	128	54	16	4
FP	2148	1563	984	514	230	80	22	5	1
TN	13322	13907	14486	14929	15240	15390	15448	15465	15469
FN	5	34	104	217	363	465	539	577	593
Operating characteristics									
Sens	0.99 (0.98-1.00)	0.94 (0.92-0.96)	0.82 (0.79-0.85)	0.63 (0.59-0.67)	0.39 (0.35-0.43)	0.22 (0.26-0.34)	0.09 (0.11-0.17)	0.03 (0.02-0.04)	0.01 (0.0-0.02)
Spec	0.86 (0.86-0.87)	0.90 (0.92-0.96)	0.94 (0.93-0.94)	0.97 (0.96-0.97)	0.99 (0.98-0.99)	0.99 (0.99-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)
PPV	0.21 (0.20-0.23)	0.26 (0.24-0.28)	0.33 (0.31-0.36)	0.41 (0.38-0.44)	0.5 (0.45-0.55)	0.62 (0.55-0.68)	0.71 (0.60-0.81)	0.76 (0.53-0.92)	0.80 (0.28-0.99)
NPV	1.0 (1.0-1.0)	1.0 (1.0-1.0)	0.99 (0.99-0.99)	0.99 (0.98-0.99)	0.98 (0.97-0.98)	0.97 (0.97-0.97)	0.97 (0.96-0.97)	0.96 (0.96-0.97)	0.96 (0.96-0.97)
PLR	7.14 (6.86-7.43)	9.33 (8.87-9.82)	13.0 (12.1-13.9)	18.1 (16.4-20.1)	26.1 (22.2-30.1)	41.7 (32.0-54.5)	64.0 (39.3-104.4)	83.5 (30.7-227.1)	104.3 (11.7-932.2)
NLR	0.01 (0.00-0.02)	0.06 (0.05-0.09)	0.19 (0.16-0.22)	0.38 (0.34-0.42)	0.62 (0.58-0.66)	0.79 (0.76-0.82)	0.91 (0.89-0.93)	0.97 (0.96-0.99)	0.99 (0.98-1.0)
DOR	704 (291-1700)	179 (117-276)	86 (67-113)	54 (44-65)	39 (32-47)	42 (33-52)	62 (42-91)	118. (48-288)	211 (26-1694)

Sens-sensitivity, Spec-specificity, PPV-positive predictive value, NPV-negative predictive value, PLR-positive likelihood ratio, NLR-negative likelihood ratio, DOR-diagnostic odds ratio
Items in brackets () 95% confidence interval

7.4 Conclusion

This chapter has described the development of the SEPSIS screening tool. Development adhered to the methods proposed by Labarère *et al*,[58] however not all elements of the checklist were reported in this chapter. First predictors of sepsis were identified, followed by the development of several multivariable models. Appropriate intervals were established for each variable, referring to existing standards, and by calculation of optimal cut points. Each of the models was compared, and the best performing model selected. Model calibration and discrimination have been reported. Operating characteristics for successive SEPSIS point scores have been calculated.

In an undifferentiated medical population, applying a cut-off of 3 or higher, sensitivity for the SEPSIS screening tool was 0.80 (95%CI 0.74-0.84), specificity was 0.78 (95%CI 0.77-0.79), positive predictive value was 0.12 (95%CI 0.10-0.14), negative predictive value was 0.99 (95%CI 0.99-0.99), positive likelihood value was 3.56 (95%CI 3.30-3.85), negative likelihood valve was 0.26 (95%CI 0.21-0.34) and the diagnostic odds ratio was 13.5 (95%CI 9.9-18.4).

If use of the SEPSIS screening tool was limited to patients with infection, applying a cut-off of 3 or higher, sensitivity for the SEPSIS screening tool was 0.82 (95%CI 0.79-0.85), specificity was 0.94 (95%CI 0.93-0.94), positive predictive value was 0.33 (95%CI 0.31-0.36), negative predictive value was 0.99 (95%CI 0.99-0.99), positive likelihood value was 13.0 (95%CI 12.1-13.9), negative likelihood valve was 0.19 (95%CI 0.16-0.22) and the diagnostic odds ratio was 86 (95%CI 67-113).

Internal validation utilising a clean dataset is reported will be **chapter 8**, while comparison of the SEPSIS screening tool with different models will be reported **chapter 9**.

Chapter 8

Validation of the SEPSIS screening tool

"Any classification system, be it nominal, ordinal, or scalar, should be proved to be a workable tool before it is used in a discriminatory or predictive manner." (A.H. Burnstein, 1993)

8.1 Introduction

There are several potential approaches to validation of a clinical prediction tool. This chapter describes the internal validation of the SEPSIS screening tool based upon a split-sampling method. An overview of approaches to validation will be provided along with a justification for the method ultimately adopted. The chapter includes a description of the data used to validate the SEPSIS screening tool and will report the operating characteristics of the SEPSIS screening tool when applied to a new dataset. In addition, model calibration, model discrimination and goodness of fit will be assessed.

8.2 Background

It is essential to determine if a model derived from one population is reliable when applied to a different population.[201] Conceptually, this is referred to as generalisability or validity. A model that generates reliable predictions in a new population is said to have been validated.[59] Statistical estimates in turn, are dependent upon adequate sample size from which the estimates are calculated.

There are no hard and fast rules dictating the sample size required to validate a clinical prediction model.[202] It has been suggested that at least ten instances of the outcome of interest are required per candidate predictor in order to ensure statistically valid results.[59 69 71 175-177] On the other hand, Vergouwe *et al*[178] advise at least 100 events and 100 non-events are required for assessing model performance. They conducted a series of simulation studies and reported that, for a logistic regression model comprising six predictors, a validation sample comprising 100 events had approximately 80% power to detect mis-calibration, with predicted probabilities 1.5 times too high (or too low) on an odds scale, and a 0.1 decrease in the c-statistic.[178] The c-statistic, also called the concordance statistic, can be used to assess goodness of fit in a logistic regression model. It is equal to the area under the receiver operating characteristics curve.[203] For the

detection of smaller differences in model performance, larger sample sizes comprising more than 250 events and 250 non-events are believed to be required.[59]

Several approaches to selecting the validation population exist, however they can be broadly categorised as either external or internal. External validation implies use of a sample from a population different to that from which it was derived, for example from another geographic region, or from a different dataset. Internal validation implies use of a sample drawn for the same population, for example by dividing the primary data sample into derivation and validation cohorts.[58 67 176 177]

In general, external validation methods are preferred to internal validation.[58] External validation methods maximise use of the primary dataset as no data are set aside for validation. Maximising the available data improves model accuracy. However, if the primary dataset is sufficiently large to generate reliable statistical estimates, there is no disadvantage to undertaking an internal validation by dividing the dataset into derivation and validation cohorts (split-sampling). An additional potential benefit of external validation is the opportunity to test the model on a population that may have a different makeup, for example ethnicity or age distribution, rather than the population on which the model was derived. This helps to confirm generalisability beyond the derivation population.

Internal validation methods broadly comprise split-sampling, cross-validation or bootstrapping methods.[58] In split-sampling the primary dataset is divided into derivation and validation subsets. Typically, the primary dataset is divided into equal halves or in a 2:1 ratio with the two thirds of the sample being used to derive the model and one third to validate the model. Split-sampling should only be undertaken if the sample size is large, as to do so with small samples reduces model accuracy as estimates have more variability, and potentially leads to imprecise performance estimates.[59 201]

Cross-validation attempts to address the potential weaknesses of spilt-sampling in small population samples.[71] The primary sample is divided into multiple subsets, for example ten equal size subsets. The model is derived on a population composed of nine of the ten samples, and validated on the remaining sample. The approach differs from split-sampling in that this process is repeated multiple times. Each time the validation sample is different, while the derivation sample

contains slightly different data each time. Performance is estimated across the repeated measures. The advantage of cross-validation over split-sampling is that cross-validation utilises a much larger proportion of the data to derive the model. In the example previously stated each repetition utilises 90% of available data for derivation while 10% is used for validation. Because of the repeated iterations utilising the different subsets, all data will eventually be used in both derivation and validation datasets. Cross-validation can yield differing statistical estimates with each iteration, consequently it may be necessary to have more than 200 replications to derive an accurate estimate.

Bootstrapping, an alternative to cross-validation, is a popular choice for validating prognostic models, particularly in samples of limited size.[58] It is a non-parametric technique, whereby alternate versions of the original dataset are created by randomly re-sampling from the original dataset.[58 59 71] For example, if the original dataset comprises the letters A, B, C and D, bootstrap samples could include:

- A, B, D, D
- A, C, D, D
- B, C, C, C
- D, D, D, D

Because the bootstrap samples are drawn from the original sample they are comparable, but not identical, to the original dataset.[68] Model performance is assessed in the original dataset and each bootstrap dataset.[204] The bootstrap datasets are plausible alternate populations with assumptions being made regarding performance of the model in plausible bootstrap populations. A statistical estimate is generated by approximating the ideal bootstrap distribution. Bootstrapping is a computationally intensive approach as many thousands of bootstrap samples are required to generate reliable results.

8.3 Methods

No external data was available, hence external validation was not possible. A single data sample was utilised to develop and validate the SEPSIS screening tool. The data sample was sufficiently large, in that it contained sufficient numbers of events and non-events in relation to the numbers of predictors. Thus there was no need to utilise cross-validation or bootstrap approaches. This study utilises an internal validation approach, using a split-sample methodology.

8.3.1 Overview of the validation dataset

Characteristics of the derivation and validation populations are reported in **table 8.1**. Prevalence of sepsis (all risk categories) is reported in **figures 8.1** and **8.2**.

Table 8.1 Population characteristics

Variable (n=6882)		Derivation	Validation
Location	*Home, n (%)*	11408 (71)	4964 (72)
	Nursing home, n (%)	1028 (6)	414 (6)
	Other, n (%)	3627 (23)	1504 (22)
	Age (years), mean (SD)	63 (21)	62 (21)
Gender	*Male, n (%)*	3346 (49)	3346 (49)
	Respirations (breaths/min), mean (SD)	20 (6)	20 (6)
	Oxygen saturation (%), mean (SD)	96 (5)	96 (5)
	Heart rate (beats/min), mean (SD)	92 (24)	92 (24)
	Systolic blood pressure (mmHg), mean (SD)	133 (27)	133 (27)
	Diastolic blood pressure (mmHg), mean (SD)	78 (17)	78 (17)
	Temperature (°C), mean (SD)	36.8 (0.9)	36.8 (0.9)
	Blood sugar (mmol/L), mean (SD)	7.0 (3.4)	7.0 (3.3)
	Glasgow Coma Score, median (IQR)	15 (15-15)	15 (15-15)
Capillary bed refill time	*Normal (<2 sec), n (%)*	13194 (82)	6567 (95)
	Delayed (>2 sec), n (%)	744 (5)	315 (5)
Skin	*Normal, n (%)*	10366 (73)	5320 (77)
	Pallor, n (%)	2669 (19)	1037 (15)
	Flushed, n (%)	856 (6)	359 (5)
	Cyanosed, n (%)	181 (1)	83 (1)
	Jaundice, n (%)	92 (0.6)	51 (0.7)
	Mottled, n (%)	53 (0.4)	20 (0.2)
	Rash, n (%)	21 (0.1)	12 (0.1)
	Pupil size (mm), median (IQR)	3 (3-4)	3 (3-4)
Pupil reaction	*Brisk, n (%)*	13447 (93)	6462 (94)
	Sluggish, n (%)	923 (6)	381 (6)
	Fixed, n (%)	128 (0.5)	39 (0.5)

Figure 8.1 Prevalence of sepsis

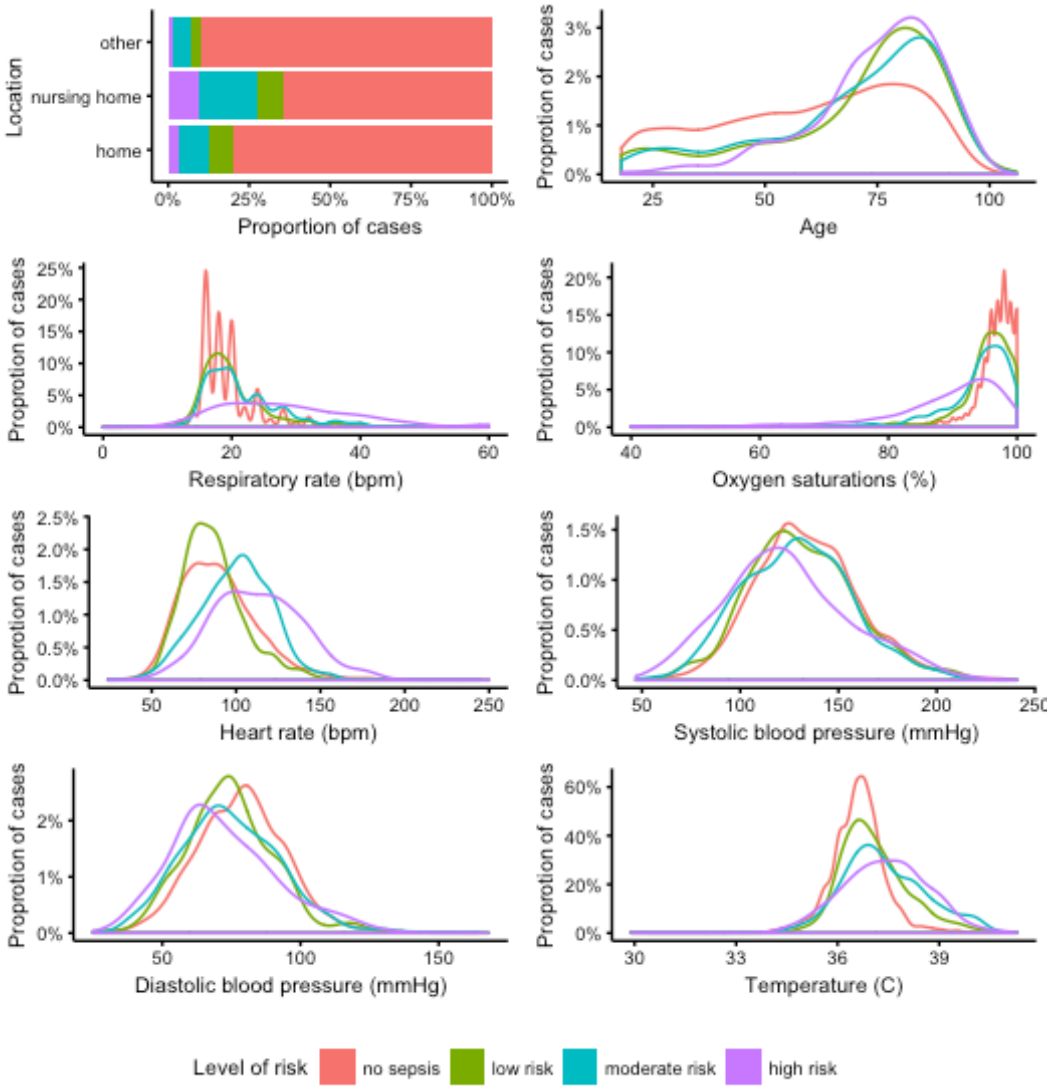
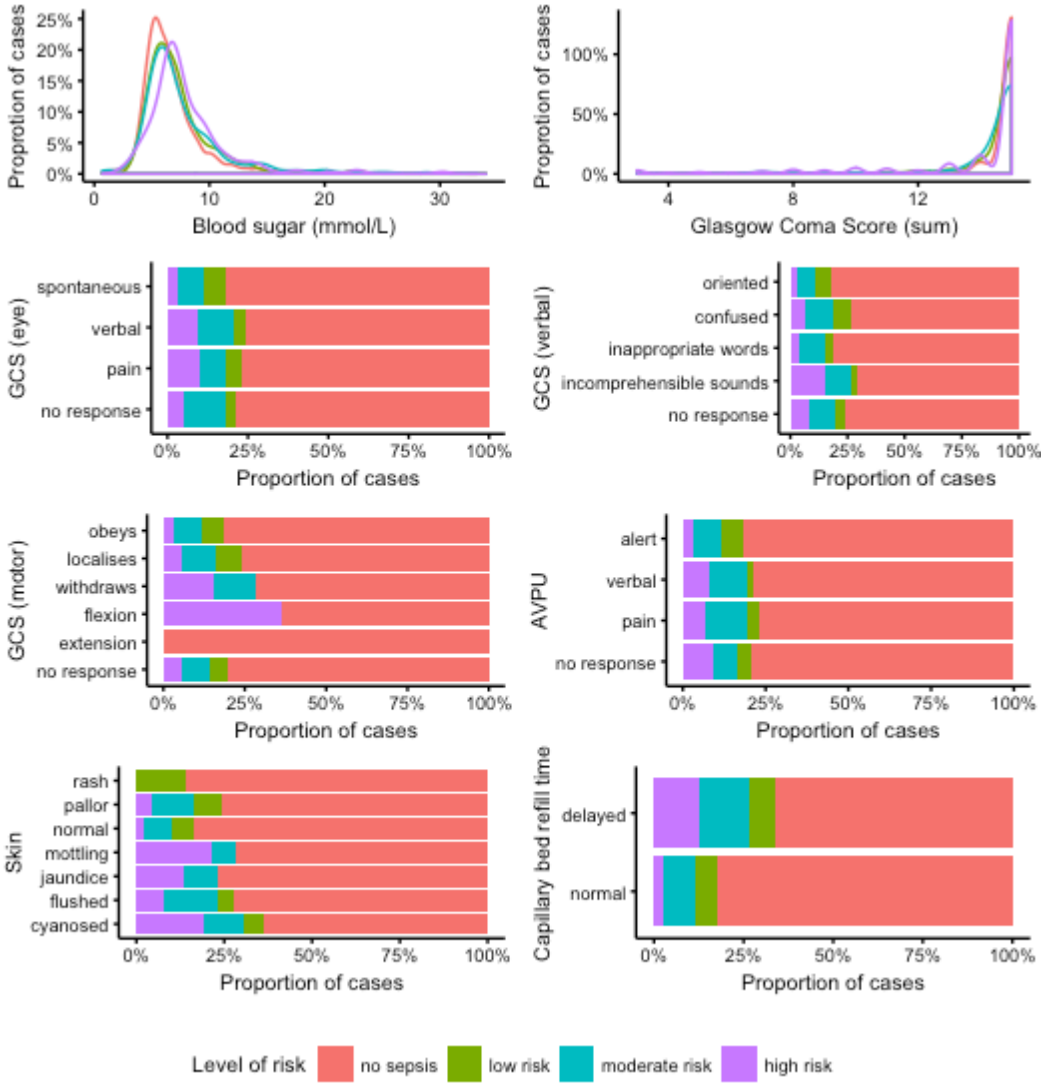


Figure 8.2 Prevalence of sepsis (continued)



8.3.2 Model calibration

Goodness of fit for the SEPSIS screening tool was assessed by calculating the Hosmer-Lemeshow statistic. A statistically significant ($p<0.05$) Hosmer-Lemeshow statistic commonly suggests that the model is overfitted, however a significant Hosmer-Lemeshow statistic is often observed when models have been derived from large samples.[191] Where the Hosmer-Lemeshow statistic is significant, suggesting the model is overfitted, assessment of the model calibration slope provides an alternate means to determine if indeed the model is overfitted.[192]

8.3.3 Model discrimination

Discrimination of the SEPSIS screening tool was assessed by calculating the area under the receiver operating characteristic curve (AUC) for the outcome of high sepsis risk at hospital. A perfect model will score "1", while a model that performs no better than chance alone will score "0.5". In the context of medical diagnosis an AUC of 0.95 or higher is preferred.

8.3.4 Brier score

Accuracy of the SEPSIS screening tool was assessed by calculating the Brier score. The Brier score represents the mean squared difference between the predicted outcome and actual outcome. The lower the Brier score the more accurate the model is. A model with perfect prediction will score "0".

8.3.5 Model operating characteristics

Performance of the SEPSIS screening tool was assessed by calculating sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, diagnostic odds ratio as well as the area under the curve for each point score.

8.4 Results

8.4.1 Overview of the validation dataset

To ensure adequate model performance and to maximise the size of the derivation data sample, the combined multiple imputed dataset was randomly partitioned (using the R package '`Caret`') into derivation and validation cohorts, such that each imputation of the validation cohort included 250 instances of high-risk sepsis. This resulted in the combined dataset being divided in a 70:30 split as reported in **table 8.2**. The number of events per predictor variable was 31 (254/8). The number of high risk cases present in the validation dataset is therefore sufficient to calculate reliable estimates.

Table 8.2 Sepsis cases in validation dataset

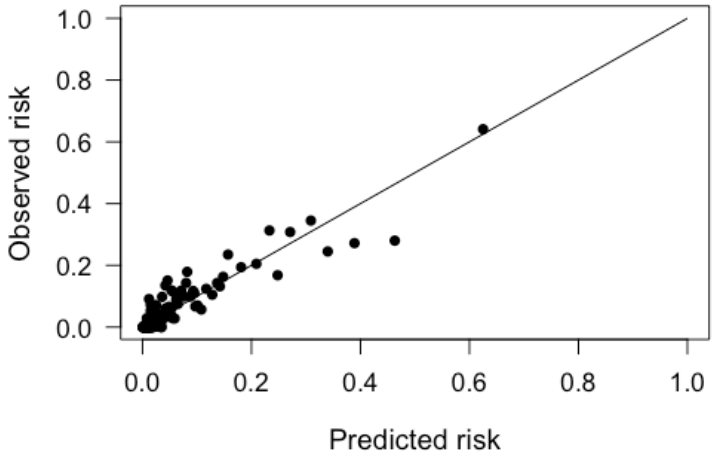
Dataset	**No sepsis risk** *n* (%)	**Low Risk Sepsis** *n* (%)	**Moderate/High Risk Sepsis** *n* (%)	**High Risk Sepsis** *n* (%)
WMAS (complete)	18690 (81.4)	1496 (6.5)	1912 (8.3)	847 (3.7)
WMAS (validation dataset)	5607 (81.5)	448 (6.5)	573 (8.3)	254 (3.7)

8.4.2 Calibration of the SEPSIS screening tool

The Hosmer-Lemeshow goodness-of-fit test was statistically significant suggesting inadequate fit (λ^2= 16.8; p<0.03). However, a statistically significant Hosmer-Lemeshow goodness-of-fit statistic is common where large data samples have been used, consequently this finding should be interpreted with caution.

Plotting observed risk versus predicted risk suggests the data are not overfitted, and indeed the calibration slope of 0.97 confirms this to be the case (see **figure 8.3**).

Figure 8.3 Calibration plot for the SEPSIS screening tool



Further evidence of adequate model specification is seen by examining the observed versus predicted outcomes (based on predicted risk) generated when calculating the Hosmer-Lemeshow statistic (see **table 8.3**).

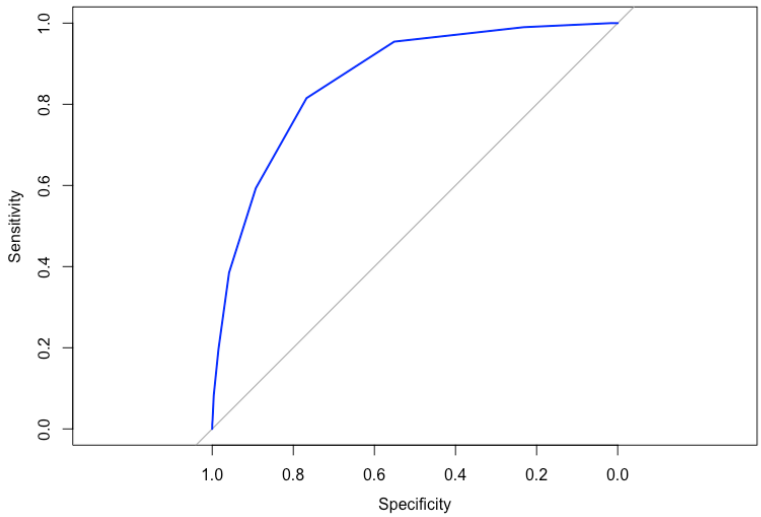
Table 8.3 Hosmer-Lemeshow observed vs expected risk

Estimated risk	Observed		Predicted	
Interval	no sepsis	sepsis	no sepsis	sepsis
0.00057,0.00285	570	0	569	1
0.00285,0.00424	567	1	566	2
0.00424,0.00651	569	0	566	3
0.00651,0.00716	613	0	609	4
0.00716,0.0104	538	4	537	5
0.0104,0.0138	540	5	539	7
0.0138,0.021	593	17	599	11
0.021,0.0378	516	18	518	16
0.0378,0.0956	508	45	518	35
0.0956,0.892	441	127	435	133

8.4.3 Model discrimination of the SEPSIS screening tool

Area under the receiver operating characteristic curve was 0.86 (95% CI 0.84-0.88) for the outcome of high sepsis risk at hospital (see **figure 8.4**). This is below the desired 0.95 that would indicate excellent diagnostic capability, but an AUC above 0.8 suggests the SEPSIS screening tool performs adequately.

Figure 8.4 Area under the Receiver Operating Characteristics (ROC) Curve



8.4.4 Brier score

Brier score for the SEPSIS screening tool was 0.0528. This value is close to "0" indicating the SEPSIS screening tool demonstrated very high levels of accuracy when evaluated with the validation dataset.

8.4.5 Operating characteristics of the SEPSIS screening tool

Several measures of model performance were calculated including sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio and diagnostic odds ratio for each point score. Performance characteristics of the SEPSIS score among adult medical patients is reported in **table 8.4**, while performance in patients with infection is reported in **table 8.5**. **Figure 8.5** graphically depicts probability of sepsis for each point on the SEPSIS score.

Table 8.4 Operating characteristics for the SEPSIS score among patients with undifferentiated medical complaints (validation data)

score	≥1	≥2	≥3	≥4	≥5	≥6	≥7	≥8	≥9
Outcome									
TP	251	241	202	145	95	47	19	5	3
FP	5028	3750	1479	671	255	91	20	6	1
TN	1600	2878	5149	5957	6373	6537	6608	6622	6627
FN	3	13	52	109	159	207	235	249	251
Operating characteristics									
Sens	0.99 (0.97-1.00)	0.95 (0.91-0.97)	0.80 (0.74-0.84)	0.57 (0.51-0.63)	0.37 (0.31-0.44)	0.19 (0.14-0.24)	0.07 (0.05-0.11)	0.02 (0.01-0.05)	0.01 (0.00-0.03)
Spec	0.24 (0.23-0.25)	0.57 (0.55-0.58)	0.78 (0.77-0.79)	0.90 (0.89-0.91)	0.96 (0.96-0.97)	0.99 (0.98-0.99)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)
PPV	0.05 (0.04-0.05)	0.08 (0.07-0.09)	0.12 (0.10-0.14)	0.18 (0.15-0.21)	0.27 (0.23-0.32)	0.34 (0.26-0.43)	0.49 (0.32-0.65)	0.45 (0.17-0.77)	0.75 (0.19-0.99)
NPV	1.00 (0.99-1.00)	1.00 (0.99-1.00)	0.99 (0.99-0.99)	0.98 (0.98-0.99)	0.98 (0.97-0.98)	0.97 (0.96-0.97)	0.97 (0.96-0.97)	0.96 (0.96-0.97)	0.96 (0.96-0.97)
PLR	1.30 (1.28-1.33)	2.19 (2.10-2.27)	3.56 (3.30-3.85)	5.6 (4.96-56.41)	9.7 (7.96-11.87)	13.5 (9.7-18.73)	24.8 (13.4-45.9)	21.75 (6.68-70.8)	78.3 (8.17-750)
NLR	0.05 (0.02-0.15)	0.09 (0.05-0.15)	0.26 (0.21-0.34)	0.48 (0.41-0.55)	0.65 (0.59-0.72)	0.83 (0.78-0.88)	0.93 (0.90-0.96)	0.98 (0.96-1.0)	0.99 (0.98-1.0)
DOR	26.6 (8.5-83.2)	24.2 (13.8-42.3)	13.5 (9.9-18.4)	11.8 (9.1-15.3)	14.9 (11.2-19.8)	16.3 (11.2-23.8)	26.7 (14.1-50.7)	22.2 (6.7-73.1)	79.2 (8.2-764.1)

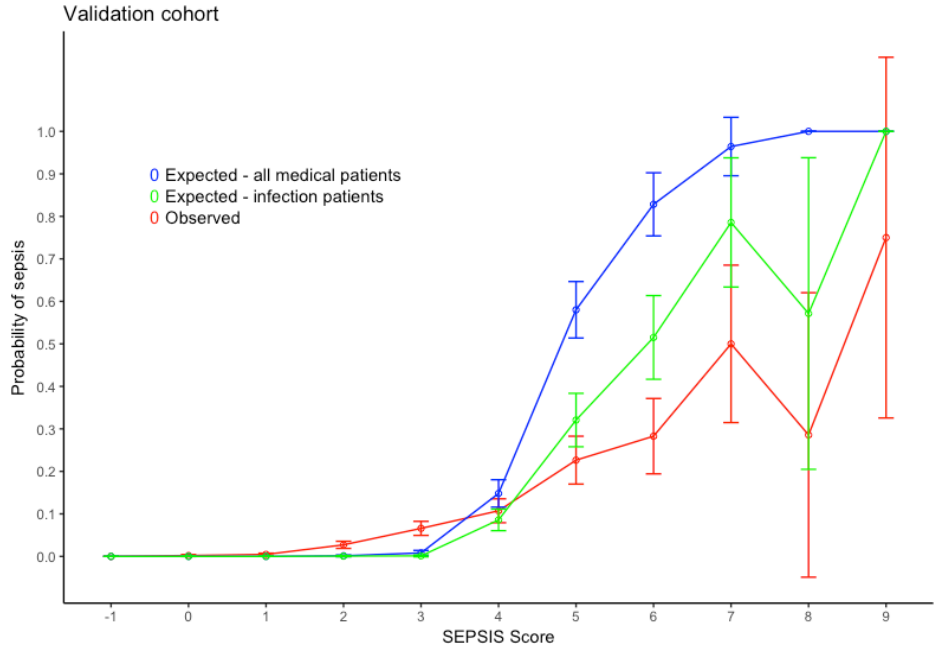
Sens-sensitivity, Spec-specificity, PPV-positive predictive value, NPV-negative predictive value, PLR-positive likelihood ratio, NLR-negative likelihood ratio, DOR-diagnostic odds ratio.
Items in brackets () 95% confidence interval

Table 8.5 Operating characteristics for the SEPSIS score among patients with infection (validation data)

score	≥1	≥2	≥3	≥4	≥5	≥6	≥7	≥8	≥9
Outcome									
TP	251	241	202	145	95	47	19	5	3
FP	933	691	1479	671	100	42	12	3	1
TN	5695	5937	5149	5957	6528	6586	6616	6625	6627
FN	3	13	52	109	159	207	235	249	251
Operating characteristics									
Sens	0.99 (0.97-1.00)	0.95 (0.91-0.97)	0.80 (0.74-0.84)	0.57 (0.51-0.63)	0.37 (0.31-0.44)	0.19 (0.14-0.24)	0.07 (0.05-0.11)	0.02(0.01-0.05)	0.01 (0.00-0.03)
Spec	0.86 (0.85-0.87)	0.90 (0.89-0.90)	0.93 (0.93-0.94)	0.96 (0.96-0.97)	0.98 (0.98-0.99)	0.99 (0.99-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)
PPV	0.21 (0.19-0.24)	0.26 (0.23-0.29)	0.32 (0.28-0.36)	0.38 (0.33-0.43)	0.49 (0.42-0.56)	0.53 (0.42-0.63)	0.61 (0.42-0.78)	0.62 (0.24-0.91)	0.75 (0.19-0.99)
NPV	1.0 (1.0-1.0)	1.0 (1.0-1.0)	0.99 (0.99-0.99)	0.98 (0.98-0.99)	0.98 (0.97-0.98)	0.97 (0.97-0.97)	0.97 (0.96-0.97)	0.96 (0.96-0.97)	0.96 (0.96-0.97)
PLR	7.02 (6.60-7.46)	9.1 (8.43-9.82)	12.2 (10.9-13.6)	16.2 (113.7-19.1)	24.8 (19.3-31.9)	29.2 (19.6-43.4)	41.3 (20.3-84.2)	43.5 (10.5-181)	78.3 (8.2-750)
NLR	0.01 (0.00-0.04)	0.06 (0.03-0.10)	0.22 (0.17-0.28)	0.44 (0.39-0.51)	0.64 (0.58-0.70)	0.82 (0.77-0.87)	0.93 (0.90-0.96)	0.98 (0.96-1.0)	0.99 (0.98-1.0)
DOR	511 (163-1598)	159 (90.7-280)	55.6 (40.4-76.5)	36.3 (27.5-48.1)	39.0 (28.3-53.8)	35.6 (23.0-55.2)	44.6 (21.4-92.9)	44.3 (10.5-187)	79.2 (8.2-764)

Sens-sensitivity, Spec-specificity, PPV-positive predictive value, NPV-negative predictive value, PLR-positive likelihood ratio, NLR-negative likelihood ratio, DOR-diagnostic odds ratio
Items in brackets () 95% confidence interval

Figure 8.5 Probability of sepsis for each point score



8.4.6 Choosing the appropriate cut-off for the SEPSIS score

A perfect sepsis screening tool would accurately discriminate between those patients who have sepsis, and those who do not. Sensitivity and specificity are the most commonly reported metrics to describe how a test performs. Sensitivity can be defined as the proportion of subjects who have the disease and are classified as disease positive by the test. Specificity can be defined as the proportion of subjects who do not have the disease and are classified as disease negative by the test. There is almost always a trade-off between sensitivity and specificity. That is, highly sensitive tests ensure few patients are missed (false negative cases), but commonly result in many false positive cases. On the other hand, highly specific tests minimise false positive cases, but commonly result in many missed cases (false negatives).

The context in which the screening tool is being used will often inform the balance between sensitivity and specificity. For example, in a disease that is rapidly fatal, a highly sensitive test is desirable to ensure patients with the disease are not missed. The clinician will tolerate a high proportion of "false negative" results, to ensure no patient with the disease "slips through the net". On the other hand, if the decision relates to a very expensive intervention, or an intervention with significant unpleasant side effects, a highly specific test may be preferred to ensure only patients with the disease are treated. As a result,

patients lacking overt features of a disease may be missed by a highly specific test, resulting in many patients "slipping through the net".

With respect to the SEPSIS score, the context of use will impact the choice of cut-off used. If the SEPSIS score is being used to identify patients at risk of sepsis, then a low cut-off score, for example SEPSIS score >=3, may be appropriate. Very few patients, who have sepsis, will be classified as not having sepsis, using a score of 3. However, using a score of 3, a significant number of patients who do not have sepsis, will be incorrectly classified as having sepsis. If the SEPSIS score was being used to decide which patients required antibiotic treatment for sepsis, then a cut-off of 3 would not be advised as this would result in many patients receiving antibiotics inappropriately. In this instance, using a higher score of 5 would significantly reduce the inappropriate administration of antibiotics, but at the expense of missing several patients who have less pronounced features of sepsis.

The objective of this thesis is to develop a screening tool to assist paramedics in identifying sepsis. It is not, in the first instance, intended to be a trigger for treatment. In addition, later chapters will compare performance of the SEPSIS score with alternate screening tools, concerning their ability to identify sepsis. As such, a pragmatic approach to balance sensitivity with specificity has been adopted in this thesis, rather than one that favours sensitivity over specificity or vice versa. In other words, rather than choose an approach that identifies all patients with sepsis (maximises sensitivity), or one that only identifies patients with sepsis (maximises specificity), I have sought to identify most patients with sepsis, while accepting that there will be false positive cases. Over-triage, as occurs in this example, is a common and accepted approach in prehospital medicine.

The ROC curve is a graphical representation of the trade-off between sensitivity and specificity and can inform the choice of cut-off adopted. **Figure 8.4** demonstrates how sensitivity and specificity are related. A SEPSIS score >=3 has been adopted as this represents a good balance between sensitivity and specificity in the undifferentiated medical population. That is, sensitivity is 0.80 indicating 80% of patients with sepsis are identified by this cut point (20% of sepsis cases are missed). Specificity is 0.78 indicating that 78% of patients without sepsis are identified as being disease free (22% are incorrectly identified

as having sepsis). This represents an acceptable over triage to minimise missing sepsis cases without overwhelming the hospital with too many false positive cases. Such over triage would not be appropriate if the SEPSIS score were to be used to guide antibiotic treatment.

8.5 Comparison of SEPSIS screening tool performance on derivation and validation datasets

Performance of a clinical prediction model may be substantially different when compared across the derivation and validation datasets. The derived model may be overly optimistic, leading to poor performance in the validation assessment. There are several potential reasons for differences in performance including overfitting during derivation, omission of important predictors and differences in the population.

When performance of the SEPSIS screening tool is compared across the derivation and validation datasets, performance estimates are comparable (see **table 8.6**).

Table 8.6 Comparison of performance across datasets (SEPSIS score≥3)

Estimate	Derivation cohort	Validation cohort
Sensitivity	0.82 (95%CI 0.79-0.85)	0.80 (95%CI 0.74-0.84)
Specificity	0.77 (95%CI 0.76-0.78)	0.78 (95%CI 0.77-0.79)
Positive predictive value	0.12 (95%CI 0.11-0.13)	0.12 (95%CI 0.10-0.14)
Negative predictive value	0.99 (95%CI 0.99-1.00)	0.99 (95%CI 0.99-0.99)
Positive likelihood ratio	3.57 (95%CI 3.41-3.75)	3.56 (95%CI 3.30-3.85)
Negative likelihood ratio	0.23 (95%CI 0.19-0.27)	0.26 (95%CI 0.21-0.34)
Diagnostic odds ratio	15.7 (95%CI 12.6-19.4)	13.5 (95%CI 9.9-18.4)
Area under ROC curve	0.87 (95% CI 0.85-0.88)	0.86 (95% CI 0.84-0.88)

ROC - Receiver operating characteristic

8.6 Conclusion

This chapter described the internal validation of the SEPSIS screening tool. Validation employed a split-sample methodology. A description of the data utilised was provided and performance of the SEPSIS screening tool with respect to goodness of fit, model calibration, model discrimination, Brier score and operating characteristics was provided.

Using a threshold of 3 points, the SEPSIS screening tool was identified to have a sensitivity of 0.80 (95%CI 0.74-0.84), a specificity of 0.78 (95%CI 0.77-0.79), a

positive predictive value of 0.12 (95%CI 0.10-0.14), a negative predictive value of 0.99 (95%CI 0.99-0.99), positive likelihood ratio was 3.56 (95%CI 3.30-3.85), a negative likelihood ratio of 0.26 (95%CI 0.21-0.34) and a diagnostic odds ratio of 13.5 (95%CI 9.9-18.4), for high risk of sepsis, among patients with an undifferentiated medical complaint.

Similarly, a threshold of 3 points had a sensitivity a sensitivity of 0.80 (95%CI 0.74-0.84), a specificity of 0.93 (95%CI 0.93-0.94), a positive predictive value of 0.32 (95%CI 0.28-0.36), a negative predictive value of 0.99 (95%CI 0.99-0.99), positive likelihood value was 12.2 (95%CI 10.9-13.6), a negative likelihood valve of 0.22 (95%CI 0.17-0.28) and a diagnostic odds ratio of 55.6 (95%CI 40.4-76.5), for high risk of sepsis, when applied to patients with infection.

Chapter 9

Comparison of prehospital sepsis screening tools

9.1 Introduction

Previous chapters have identified the need for a prehospital sepsis screening tool, and subsequently described how such a tool (the SEPSIS screening tool) was developed and validated. In this chapter, performance of the SEPSIS screening tool is assessed relative to existing screening tools. It should be noted that no prehospital sepsis screening tool has been validated in routine prehospital clinical practice.

9.2 Background

In-hospital data indicate that early administration of antibiotics in cases of severe sepsis is associated with reduced mortality.[205 206] However, the diagnosis of sepsis is often only confirmed many hours after hospital admission, when all the required clinical evidence, such as biomarkers for organ failure and blood culture results, become available.[4 15 22] Consequently, pragmatic approaches that advocate early administration of antibiotics, before a definitive diagnosis of sepsis has been made, have been developed.[4 5 15 22] The 2016 National Institute for Health Care Excellence (NICE) guideline (NG51) "Sepsis: recognition, diagnosis and early management" is one such example of a pragmatic guideline that advocates administration of antibiotics to patients with a high risk of sepsis, before a definitive diagnosis has been confirmed by laboratory tests.[21] This guideline includes an algorithm for use in acute health care settings such as the Emergency Department (ED).

It has been reported that more than two-thirds of severe sepsis cases are initially seen in the ED[173] and around half of ED sepsis patients arrive by ambulance.[26-28 170-172] Sepsis patients arriving at the ED via Emergency Medical Services (EMS) are likely to be sicker than those arriving by other means;[28 29 170-172] and up to 80% of severe sepsis patients admitted to intensive care from the ED will have been transported by EMS.[26 30]

It has been argued that there is an opportunity for ambulance clinicians to improve outcomes for this population in the same manner as they do with other time critical, life-threatening conditions such as acute myocardial infarction,[89] stroke[82] and major trauma.[85] Early recognition of sepsis by ambulance clinicians could reduce time to delivery of key interventions such as antibiotic therapy. However, evidence suggests that recognition of sepsis by paramedics is often poor.[31 32 36 108 170] Use of a prehospital sepsis screening tool to optimise prehospital recognition of sepsis has been advocated.[40 173 207] A prehospital sepsis screening tool, able to reliably identify those patients who are categorised as at high risk of sepsis as per NICE guideline (NG51) "Sepsis: recognition, diagnosis and early management", could reduce the interval to administration of antibiotics, and potentially improve outcomes among patients with sepsis.[21]

9.3 Method

To enable a true comparison of screening tool performance, the same dataset utilised to validate the SEPSIS screening tool, is used to assess performance of the alternate screening tools. This dataset is comprised of 6882 adult patients, attended by West Midlands Ambulance Service NHS Foundation Trust (WMAS), and transported to University Hospital North Staffordshire NHS Foundation Trust (UHNS) between 01 July 2013 and 30 June 2014. Patients under 18 years of age, and all cases of cardiac arrest, trauma or mental health diagnosis were excluded. The included population therefore comprised adult patients with a medical complaint only. The gold standard hospital diagnosis is *"high risk of sepsis"* as defined by the 2016 National Institute for Health Care Excellence (NICE) guideline (NG51) "Sepsis: recognition, diagnosis and early management". Diagnostic criteria for sepsis are outlined in **table 9.1**. High risk sepsis status was determined by examination of hospital (UHNS) clinical data, and not extracted from diagnostic fields in the hospital record.

Table 9.1 Classification of risk of sepsis (NICE 2016)

Category	High risk criteria	Moderate to high risk criteria	Low risk criteria
History	Objective evidence of new altered mental state	History from patient, friend or relative of new onset of altered behaviour or mental state History of acute deterioration of functional ability Impaired immune system (illness or drugs including oral steroids) Trauma, surgery or invasive procedures in the last 6 weeks	Normal behaviour
Respiratory	Raised respiratory rate: 25 breaths per minute or more New need for oxygen (more than 40% FiO2) to maintain saturation more than 92% (or more than 88% in known chronic obstructive pulmonary disease)	Raised respiratory rate: 21–24 breaths per minute	No high risk or moderate to high risk criteria met
Blood pressure	Systolic blood pressure 90 mmHg or less or systolic blood pressure more than 40 mmHg below normal	Systolic blood pressure 91–100 mmHg	No high risk or moderate to high risk criteria met
Circulation and hydration	Raised heart rate: more than 130 beats per minute Not passed urine in previous 18 hours. For catheterised patients, passed less than 0.5 ml/kg of urine per hour	Raised heart rate: 91–130 beats per minute (for pregnant women 100–130 beats per minute) or new onset arrhythmia Not passed urine in the past 12–18 hours For catheterised patients, passed 0.5–1 ml/kg of urine per hour	No high risk or moderate to high risk criteria met
Temperature		Tympanic temperature less than 36°C	
Skin	Mottled or ashen appearance Cyanosis of skin, lips or tongue Non-blanching rash of skin	Signs of potential infection, including redness, swelling or discharge at surgical site or breakdown of wound	No non-blanching rash

Performance of the new "screening to enhance prehospital identification of sepsis (SEPSIS) tool", and existing tools "out of hospital NICE sepsis risk", "National Early Warning Score (NEWS)", "Modified Early Warning Score (MEWS)", "prehospital early sepsis detection score (PRESS)", "prehospital identification of severe sepsis (PreSep)", "critical illness score (CIS)", "UK Sepsis Trust screening tool (Red Flag Sepsis)", "Swedish '*Andningsfrekvens*' (BAS 90-30-90)", Survive Sepsis Campaign "quick Sepsis Organ Failure Assessment (qSOFA)" and the 2012 Survive Sepsis Campaign definition utilising two or more systemic inflammatory response syndrome criteria (SIRS) were compared.

The SEPSIS screening tool has been described in detail in previous chapters. The 2016 NICE guideline (NG51) "Sepsis: recognition, diagnosis and early management" includes an algorithm specifically for use outside the acute hospital environment, that ambulance clinicians would use. The algorithm requires the user to identify those patients at risk of sepsis by considering a likely source of infection as well as additional risk factors and indicators of clinical concern. Patients who are at risk of sepsis are then stratified into high, moderate and low risk categories. Indicators of high risk include objective evidence of new altered mental state, respiratory rate greater than 25 breaths/min (or the need for supplemental oxygen to maintain normal oxygen saturations), heart rate of 130 beats/min or higher, systolic blood pressure of 90 mmHg or lower, not passed urine for 18 hours (or passed less than 0.5mlkg if catheterised), mottled or ashen appearance, cyanosis of skin, lips or tongue or presence of a non-blanching rash. A patient at risk of sepsis with any one high risk indicator is classified as being at high risk of sepsis.[21]

NEWS is an early warning score, developed by the Royal College of Physicians.[208] It is intended to be used as a surveillance system for all patients in hospital, tracking their clinical condition, alerting the clinical team to any medical deterioration and triggering a timely clinical response. The NEWS score comprises six routine physiological measurements - respiratory rate, oxygen saturations, temperature, systolic blood pressure, pulse rate and level of consciousness. A score is allocated for each as they are measured, the magnitude of the score reflecting how extreme the parameter varies from the norm. This score is then aggregated, and uplifted for people requiring oxygen.[208] The

maximum NEWS score is 20 points. Although not developed specifically for sepsis screening, it has been studied in this context.[209 210]

MEWS is another early warning score, intended to detect early deterioration of surgical patients, comprising four physiological findings and one observation.[110] A MEWS score is calculated from systolic blood pressure, heart rate, respiratory rate, temperature and an assessment of consciousness using the AVPU score. Like NEWS, MEWS allocates a score for each variable that reflects deviation from the norm. The maximum MEWS score is 14 points. Although MEWS was not developed specifically for sepsis screening, its use in this context has been reported.[36 54 211]

The PRESS score was developed by Polito *et al*[108] to identify severe sepsis in adult patients. Screening follows several stages. First patients must meet the following criteria: over 18 years of age, systolic blood pressure below 110 mmHg, heart rate above 90 beats/min and respiratory rate above 20 breaths per minute. Patients not meeting these criteria are not eligible. Next, specific case types are excluded. Traumatic injury, cardiac arrest, pregnancy, psychiatric emergency or confirmed toxic ingestion excludes patients from further screening. All remaining patients are classified as at-risk and a score is calculated. Scores are allocated to six components: a dispatch code of "sick person", nursing home residence, age, hot tactile temperature, systolic blood pressure and oxygen saturations. The maximum score is 24 points. Any at-risk patient scoring above 1 is classified as having severe sepsis, however a higher points score does not indicate increasing sepsis severity.

The PreSep score was developed by Bayer *et al*[54] to identify severe sepsis in adult patients. Scores are allocated to five components: temperature above 38°C or below 36°C, oxygen saturations below 92%, respiratory rate above 22 breaths/min, heart rate above 90 beats/min and systolic blood pressure below 90mmHg. The maximum score is 11 points, and any patient scoring above 3 is classified as having severe sepsis.

The CIS score was developed by Seymour *at al*[107] to identify patients in need critical care services. Critical illness was defined as in-hospital diagnosis of severe sepsis, delivery of mechanical ventilation, or death at any point during hospitalisation. Scores are allocated to five components: age, respiratory rate, systolic blood pressure, heart rate, oxygen saturations and Glasgow Coma Score.

The maximum score is 8 points, and any patient scoring above 3 is classified as requiring critical care services. The authors describe using this score as a proxy for a diagnosis of severe sepsis requiring admission to intensive care (ICU)

"Red Flag Sepsis" is a screening algorithm developed and promoted by the UK Sepsis Trust to stratify patients according to need for senior review and initial treatment.[5] It comprises several stages, the first of which is to identify patients at risk with reference to their NEWS score and appearance. The next stage is to consider if the cause could be due to infection, and if so the final stage is to categorise the patient as either "Red Flag" (high risk of sepsis) or "Amber Flag" (moderate risk of sepsis). There are nine potential "Red Flag" indicators: responds only to voice or pain/unresponsive, systolic blood pressure 90 mmHg or lower, heart rate above 130 beats/min, respiratory rate above 24 breaths/min, supplemental oxygen required to maintain peripheral oxygen saturations of 92% or higher, presence of a non-blanching rash, mottled, ashen or cyanotic skin, not passed urine in the preceding 18 hours, lactate of 2mmol/L or higher and a recent history of chemotherapy. The presence of any one "Red Flag" implies the patient is at high risk of sepsis, that a treatment package should be implemented without delay and that the patient should be reviewed by a senior clinician as soon as possible.[5]

The BAS 90-30-90 score is a Swedish screening tool ('Andningsfrekvens') comprising systolic blood pressure below 90mmHg, respiratory rate above 30 breaths/min and oxygen saturation below 90%. The maximum score is 3 points, and any patient scoring above 0 is classified as having severe sepsis.[36]

The qSOFA score (also known as quickSOFA) is a bedside assessment to identify patients with suspected infection who are at greater risk for a poor outcome. It was developed by the Survive Sepsis Campaign and is intended for use outside the intensive care unit (ICU) environment.[4] Scores are allocated to three components: systolic blood pressure of 100mmHg or lower, respiratory rate of 22 breaths/min or higher and Glasgow Coma Scale below 15. The maximum score is 3 points, and any patient scoring 2 or 3 requires that clinicians investigate for organ dysfunction, initiate or escalate therapy as appropriate, and consider referral to critical care or increase the frequency of monitoring, if such actions have not already been undertaken.[4]

A perfect sepsis screening tool would correctly identify those patients who are septic (have sepsis), and those who are not (no sepsis). To date, no such screening tool has been identified for sepsis. Several metrics are used to describe the accuracy of screening tools, and the context in which the screening tool is being used may inform which metric is most important in the given circumstance. For example, for a disease that is rapidly fatal, a highly sensitive test may be required to ensure patients with the disease are not missed. On the other hand, if the decision relates to a very expensive intervention, or an intervention with significant unpleasant side effects, a highly specific test may be preferred to ensure only patients with the disease are treated. However, for the bedside clinician it is also vital to be able to interpret a test result and apply it to their clinical practice. In other words, if the screening tool suggests the patient has the disease, how likely is it that the patient has disease and should the clinician be more confident to provide treatment appropriate for the disease, based upon the test result? For this, positive predictive values (PPV) and positive likelihood ratios (PLR) which are defined below, are important. Conversely, if screening suggests the patient does not have the disease, should the clinician be confident that patient does not have the disease and not provide treatment for the disease? Negative predictive values (NPV) and negative likelihood ratios (NLR) which are defined below, will inform the clinician in this instance. In this chapter, screening tool performance is assessed by reporting the following measures:

- Sensitivity - the ability of a test to identify those with the disease. A test with sensitivity = 1 identifies all cases with the disease.
- Specificity - the ability of a test to identify those who do not have the disease. A test with specificity = 1 identifies all cases without the disease.
- Positive predictive value (PPV) - probability that an individual with a positive test has the disease. A test with PPV = 1 implies all cases classified as having disease will have the disease.
- Negative predictive value (NPV) - probability that an individual with a negative test does not have the disease. A test with NPV = 1 implies all cases classified as being disease free will not have the disease.
- Positive likelihood ratio (PLR) - the probability that an individual with the disease will test positive, divided by, the probability that an individual without the disease will test positive. A PLR provides an estimate of how much a positive test result will increase the odds of

having the disease (See **table 9.2** for interpretation of likelihood ratios). To interpret a PLR it is essential to know the pre-test odds of having a disease. The pre-test odds of having the disease are related to the prevalence of disease in the population being studied. In the population examined the pre-test odds for high risk sepsis is 0.037 (see **table 9.3**).

- Negative likelihood ratio (NLR) - the probability that an individual with the disease will test negative, divided by, the probability that an individual without the disease will test negative. A NLR provides an estimate of how much a negative test result will decrease the odds of having the disease (See **table 9.2** for interpretation of likelihood ratios). As with PLR, interpretation of a NLR requires knowledge of the pre-test odds of having a disease (see **table 9.3**).

Table 9.2 Interpretation of Likelihood Ratios (LR)

LR	Interpretation
> 10	Large and often conclusive increase in the likelihood of disease
5 - 10	Moderate increase in the likelihood of disease
2 - 5	Small increase in the likelihood of disease
1 - 2	Minimal increase in the likelihood of disease
1	No change in the likelihood of disease
0.5 - 1.0	Minimal decrease in the likelihood of disease
0.2 - 0.5	Small decrease in the likelihood of disease
0.1 - 0.2	Moderate decrease in the likelihood of disease
< 0.1	Large and often conclusive decrease in the likelihood of disease

9.4 Results

Characteristics of the population are reported in **table 9.3**.

Table 9.3 Validation cohort characteristics

Variable (n=6882)		Estimate
Location	*Home, n (%)*	4964 (72)
	Nursing home, n (%)	414 (6)
	Other, n (%)	1504 (22)
	Age (years), mean (SD)	62 (21)
Gender	*Male, n (%)*	3346 (49)
	Respirations (breaths/min), mean (SD)	20 (6)
	Oxygen saturation (%), mean (SD)	96 (5)
	Heart rate (beats/min), mean (SD)	92 (24)
	Systolic blood pressure (mmHg), mean (SD)	133 (27)
	Diastolic blood pressure (mmHg), mean (SD)	78 (17)
	Temperature (°C), mean (SD)	36.8 (0.9)
	Blood sugar (mmol/L), mean (SD)	7.0 (3.3)
	Glasgow Coma Score, median (IQR)	15 (15-15)
Capillary bed refill time	*Normal (<2 sec), n (%)*	6567 (95)
	Delayed (>2 sec), n (%)	315 (5)
Skin	*Normal, n (%)*	5320 (77)
	Pallor, n (%)	1037 (15)
	Flushed, n (%)	359 (5)
	Cyanosed, n (%)	83 (1)
	Jaundice, n (%)	51 (0.7)
	Mottled, n (%)	20 (0.2)
	Rash, n (%)	12 (0.1)
	Pupil size (mm), median (IQR)	3 (3-4)
Pupil reaction	*Brisk, n (%)*	6462 (94)
	Sluggish, n (%)	381 (6)
	Fixed, n (%)	39 (0.5)
NICE risk of sepsis	*No risk (no infection), n (%)*	5607 (81.5)
	Low risk, n (%)	448 (6.5)
	Moderate risk, n (%)	573 (8.3)
	High risk, n (%)	254 (3.7)

Two sets of performance metrics for each screening tool are presented. In **table 9.4** metrics are reported when applying the screening tools to an undifferentiated adult medical patient population, that is patients with all medical conditions are included. In **table 9.**5 performance metrics are reported when applying the screening tools to a population of patients with infection. In other words, **table 9.5** provides and approximation of performance of the screening tools applied to patients with suspected infection.

The most sensitive screening tool, applied to an undifferentiated adult medical population, is NEWS>=2. This screening tool maximises the number of patients correctly identified as having sepsis. However, it also generates the highest number of false positive cases. Positive predictive value is lowest for this screening tool suggesting that a clinician using this screening tool will have the least confidence that a patient with positive test result does in fact have sepsis (one patient in every 20 with a positive result will have high risk sepsis). Conversely, if the screening tool indicates that a patient does not have sepsis the treating clinician can be very confident that the patient does not have sepsis.

The most specific screening tool is the PRESS score. This screening tool maximises the number of patients correctly identified as not having sepsis. However, it also generates the highest number of false negative cases - in other words, of all the screening tools assessed this screening tool misses the most sepsis cases. PRESS fails to identify 80% of cases with high risk sepsis.

The tool with the greatest PPV is the CIS score. In an undifferentiated population, approximately one patient out of every five who test positive will have high risk sepsis, whereas in a population of patients with infection this will improve to one patient in every two with a positive result. Whilst this is very promising, the CIS score fails to identify almost 60% of high sepsis risk cases.

Table 9.4 Comparison of screening tools (All medical patients)

Screening Tool	TP FN	FP TN	Apparent prevalence	Sensitivity	Specificity	PPV	NPV	PLR	NLR
SEPSIS>=2	241 13	2878 3750	0.45 (0.44 – 0.47)	0.95 (0.74 – 0.84)	0.57 (0.77 – 0.79)	0.08 (0.07 – 0.09)	0.99 (0.99 – 1.00)	2.19 (2.10 – 2.27)	0.09 (0.05 – 0.15)
SEPSIS>=3	202 52	1479 5149	0.24 (0.23 – 0.25)	0.80 (0.74 – 0.84)	0.78 (0.77 – 0.79)	0.12 (0.10 – 0.14)	0.99 (0.99 – 0.99)	3.56 (3.30 – 3.85)	0.26 (0.21 – 0.34)
SEPSIS>=5	95 159	255 6373	0.05 (0.05 – 0.06)	0.37 (0.31 – 0.44)	0.96 (0.96 – 0.97)	0.27 (0.23 – 0.32)	0.97 (0.96 – 0.97)	24.8 (13.4 – 45.9)	0.93 (0.90 – 0.96)
CIS	114 140	398 6230	0.07 (0.07 – 0.08)	0.45 (0.39 – 0.51)	0.94 (0.93 – 0.95)	0.22 (0.19 – 0.26)	0.98 (0.97 – 0.98)	7.47 (6.33 – 8.83)	0.59 (0.52 – 0.66)
PreSep	155 99	836 5792	0.14 (0.15 – 0.15)	0.61 (0.55 – 0.67)	0.87 (0.87 – 0.88)	0.16 (0.13 – 0.18)	0.98 (0.98 – 0.99)	4.84 (4.30 – 5.44)	0.45 (0.38 – 0.52)
PRESS	45 203	170 6422	0.03 (0.03 – 0.04)	0.18 (0.14 – 0.24)	0.97 (0.97 – 0.98)	0.21 (0.16 – 0.27)	0.97 (0.96 – 0.97)	7.04 (5.20 – 9.53)	0.84 (0.79 – 0.89)
qSOFA	74 180	464 6164	0.08 (0.07 – 0.08)	0.29 (0.24 – 0.35)	0.93 (0.92 – 0.94)	0.14 (0.11 – 0.17)	0.97 (0.97 – 0.98)	4.16 (3.37 – 5.14)	0.76 (0.70 – 0.82)
SIRS>=2	204 50	2242 4385	0.36 (0.34 – 0.37)	0.80 (0.75 – 0.85)	0.66 (0.65 – 0.67)	0.08 (0.07 – 0.10)	0.99 (0.99 – 0.99)	2.37 (2.21 – 2.55)	0.30 (0.23 – 0.38)
BAS 90-30-90	159 95	847 5781	0.15 (0.14 – 0.15)	0.63 (0.56 – 0.69)	0.87 (0.86 – 0.88)	0.16 (0.14 – 0.18)	0.98 (0.98 – 0.99)	4.90 (4.37 – 5.49)	0.43 (0.37 – 0.50)
MEWS>=4	161 93	988 5640	0.17 (0.16 – 0.18)	0.63 (0.57 – 0.69)	0.85 (0.84 – 0.86)	0.14 (0.12 – 0.16)	0.98 (0.98 – 0.99)	4.25 (3.81 – 4.75)	0.43 (0.37 – 0.51)
NEWS>=2	251 3	5146 1482	0.78 (0.77 – 0.79)	0.99 (0.97 – 1.00)	0.22 (0.21 – 0.23)	0.05 (0.04 – 0.05)	1.00 (0.99 – 1.00)	1.27 (1.25 – 1.30)	0.05 (0.02 – 0.16)
NEWS>=3	247 7	3656 2972	0.57 (0.56 – 0.58)	0.97 (0.94 – 0.99)	0.45 (0.44 – 0.45)	0.06 (0.06 – 0.07)	1.00 (1.00 – 1.00)	1.76 (1.71 – 1.82)	0.06 (0.03 – 0.13)
NEWS>=5	217 37	1920 4708	0.31 (0.30 – 0.32)	0.85 (0.80 – 0.90)	0.71 (0.70 – 0.72)	0.10 (0.09 – 0.12)	0.99 (0.99 – 0.99)	2.95 (2.77 – 3.14)	0.21 (0.15 – 0.28)

Table 9.5 Comparison of screening tools (Infection patients only)

Screening Tool	TP FN	FP TN	Apparent prevalence	Sensitivity	Specificity	PPV	NPV	PLR	NLR
SEPSIS>=2	241 13	691 5937	0.14 (0.13 – 0.14)	0.95 (0.91 – 0.97)	0.90 (0.89 – 0.90)	0.26 (0.23 – 0.29)	1.00 (1.00 – 1.00)	9.10 (8.43 – 9.82)	0.06 (0.03 – 0.10)
SEPSIS>=3	202 52	433 6195	0.09 (0.09 – 0.10)	0.80 (0.74 – 0.84)	0.93 (0.93 – 0.94)	0.32 (0.28 – 0.36)	0.99 (0.99 – 0.99)	12.17 (10.90 – 13.59)	0.22 (0.17 – 0.28)
SEPSIS>=5	95 159	100 6528	0.04 (0.03 – 0.04)	0.37 (0.31 – 0.44)	0.98 (0.98 – 0.99)	0.49 (0.42 – 0.56)	0.98 (0.97 – 0.98)	24.8 (19.3 -31.9)	0.64 (0.58 – 0.70)
NICE (high risk only)	214 40	510 6118	0.11 (0.10 – 0.11)	0.84 (0.79 – 0.89)	0.92 (0.92 – 0.93)	0.30 (0.26 – 0.33)	0.99 (0.99 – 1.00)	10.95 (9.92 – 12.09)	0.17 (0.163 – 0.23)
Red Flag Sepsis (UK sepsis Trust)	212 42	449 6179	0.10 (0.09 – 0.10)	0.83 (0.78 – 0.88)	0.93 (0.93 – 0.94)	0.32 (0.29 – 0.36)	0.99 (0.99 – 1.00)	12.32 (11.10 – 13.68)	0.18 (0.13 – 0.23)
CIS	114 140	115 6513	0.03 (0.03 – 0.04)	0.45 (0.39 – 0.51)	0.98 (0.98 – 0.99)	0.50 (0.43 – 0.56)	0.98 (0.98 – 0.98)	25.87 (20.62 – 32.45)	0.56 (0.50 – 0.63)
PreSep	155 99	317 6311	0.07 (0.06 – 0.07)	0.61 (0.55 – 0.67)	0.95 (0.95 – 0.96)	0.33 (0.29 – 0.37)	0.98 (0.98 – 0.99)	12.76 (11.03 – 14.76)	0.41 (0.35 – 0.48)
PRESS	45 203	53 6569	0.01 (0.01 – 0.02)	0.19 (0.14 – 0.24)	0.99 (0.99 – 0.99)	0.46 (0.36 – 0.56)	0.97 (0.97 – 0.97)	22.75 (15.68 – 33.00)	0.82 (0.77 – 0.87)
qSOFA	74 180	125 6503	0.03 (0.03 – 0.03)	0.29 (0.24 – 0.35)	0.98 (0.98 – 0.98)	0.37 (0.30 – 0.44)	0.97 (0.97 – 0.98)	15.45 (11.93 – 20.01)	0.72 (0.67 – 0.78)
SIRS>=2	204 50	472 6156	0.10 (0.09 – 0.11)	0.80 (0.75 – 0.85)	0.93 (0.92 – 0.93)	0.30 (0.27 – 0.34)	0.99 (0.99 – 0.99)	11.28 (10.14 – 12.54)	0.21 (0.17 – 0.27)
BAS 90-30-90	159 95	216 6412	0.05 (0.05 – 0.06)	0.63 (0.56 – 0.69)	0.97 (0.96 – 0.97)	0.42 (0.37 – 0.48)	0.99 (0.98 – 0.99)	19.21 (16.34 – 22.59)	0.39 (0.33 – 0.45)
MEWS>=4	161 93	246 6382	0.06 (0.05 – 0.06)	0.63 (0.57 – 0.69)	0.96 (0.96 – 0.97)	0.40 (0.35 – 0.44)	0.99 (0.98 – 0.99)	17.08 (14.64 – 19.93)	0.38 (0.32 – 0.45)
NEWS>=2	251 3	890 5738	0.17 (0.16 – 0.17)	0.99 (0.97 – 1.00)	0.87 (0.86 – 0.87)	0.22 (0.20 – 0.25)	1.00 (1.00 – 1.00)	7.36 (6.91 – 7.83)	0.01 (0.00 – 0.04)
NEWS>=3	247 7	714 5914	0.14 (0.13 – 0.15)	0.97 (0.94 – 0.99)	0.89 (0.88 – 0.90)	0.26 (0.23 – 0.29)	1.00 (1.00 – 1.00)	9.03 (8.40 – 9.70)	0.03 (0.01 – 0.06)
NEWS>=5	217 37	472 6156	0.10 (0.09 – 0.11)	0.85 (0.80 – 0.90)	0.93 (0.92 – 0.93)	0.31 (0.28 – 0.35)	0.99 (0.99 – 1.00)	12.00 (10.85 – 13.27)	0.16 (0.12 – 0.21)

9.5 Discussion

From the metrics reported, no single tool can be deemed to be 'the best' for use by ambulance clinicians. Sensitivity needs to be balanced against specificity as well as predictive values and likelihood ratios.

In this regard, three screening tools, SEPSIS>=3, SIRS>=2 and NEWS>=5, appear to offer a respectable balance between sensitivity, specificity and PPV in an undifferentiated adult medical population. Respectively, these three screening tools, missed 52, 50 and 37 patients with a high risk of sepsis. On the other hand, the number of false positive cases, that is patients identified as high risk sepsis when they are not, are 1479, 2242 and 1920 respectively.

There is a trade-off between missing high risk of sepsis cases and increasing false positive cases. In this comparison, minimising missed cases (NEWS>=5 misses 37 cases) will results in an addition 441 false positive cases (1920-1479), whereas minimising false positive cases (SEPSIS>=3 results in 1479 false positive cases) results in an additional 15 missed high risk sepsis cases (52-37). The decision as to which is better, 15 missed cases or an increase of 441 false positive cases, is a decision needing input from multiple stakeholders within the emergency care network.

From the ambulance clinicians' perspective, practitioners will want to minimise the number of high risk sepsis patients who are missed. However, they also need to be confident in the result of the screening. PPV is marginally better for SEPSIS>=3 than for SIRS>=2 and NEWS>=5 at 0.12 (95%CI 0.10-0.14) compared with 0.08 (95%CI 0.07-0.10) and 0.10 (95% 0.09-0.12) respectively. However, all three screening tools have low PPV. This result indicates that a patient with positive test result using SEPSIS>=3 is marginally more likely to have high risk sepsis than if the test was positive using SIRS>=2 or NEWS>=5. There is a similar increase in PLR for SEPSIS>=3 than for SIRS>=2 and NEWS>=5 at 3.56 (95%CI 3.30-3.85) compared with 2.37 (95%CI 2.21-2.55) and 2.95 (95% 2.77-3.14) respectively. This indicates that an ambulance clinician with a positive test result using SEPSIS>=3 should be more confident that the high-risk sepsis categorisation is due to sepsis, rather than another cause, than if the test was positive using SIRS>=2 or NEWS>=5. However, in practice this improved

performance equates to between 2 and 4 more cases of sepsis per 100 cases screened.

9.6 Conclusion

In this chapter performance of the SEPSIS screening tool was assessed relative to existing screening tools. In an undifferentiated adult medical population, NEWS>=2 was the most sensitive, whereas PRESS was the most specific. When considering both sensitivity and specificity, three screening tools, SEPSIS>=3, SIRS>=2 and NEWS>=5, appeared to offer similar performance. PPV and PLR was greatest for SEPSIS>=3. Choosing the most appropriate screening tool to use to identify high risk sepsis patients is complex, and requires input from multiple stakeholders in the emergency care network.

Chapter 10

Conclusion

10.1 Introduction

Chapters 1 to 3 of this thesis identified a need for a reliable prehospital sepsis screening tool. Chapter 4 described the acquisition of data from a NHS ambulance service and a NHS Emergency Department (ED). This was followed by reporting the methods for linking the two datasets (chapter 5) and the management of missing data (chapter 6). Chapters 7 and 8 provided a detailed exposition of the derivation and validation of a new prehospital sepsis screening tool. Chapter 9 compared the performance of this new prehospital sepsis screening tool with existing screening tools. This concluding chapter will consider the implications of this work in the context of current prehospital clinical practice in the United Kingdom (UK).

10.2 Summary of findings

The overall incidence of high risk of sepsis, as per the National Institute for Health and Care Excellence (NICE) sepsis guideline [NG51], among adult patients in a UK ambulance service, was determined to be 2.8%, calculated as:

$$\left({}^{847\ cases\ of\ high\ sepsis\ risk}/_{33289\ unique\ cases\ -\ 2536\ child\ cases}\right) * 100$$

In this study, predictors of sepsis in the prehospital environment comprise demographic (age), Systemic Inflammatory Response Syndrome (SIRS) criteria (respiratory rate, heart rate and temperature) as well potential indicators of organ failure (peripheral oxygen saturations, systolic blood pressure and level of consciousness).

Recognition of sepsis can be challenging. Historically, a sepsis diagnosis was made combining both Systemic Inflammatory Response Syndrome (SIRS) criteria and indicators of organ failure in patients with infection.[15] However, SIRS criteria have recently fallen out of favour, as they are overly sensitive, without adequate specificity.[16 18 212] International diagnostic standards have migrated toward more quantifiable, and more specific, indicators of organ failure.[4]

Sadly, the technology to measure these specific indicators of organ failure has not yet migrated into routine, frontline ambulance use. Ambulance clinicians still have the challenging task of recognising a difficult to diagnose condition, with sub-optimal indicators of the condition. Despite these limitations, a pragmatic prehospital sepsis screening tool has been developed using clinical data from both the ambulance service and the ED.

Ambulance staff routinely attend patients with undifferentiated medical complaints. The SEPSIS screening tool has a sensitivity of 0.80 (95% CI 0.74-0.84), specificity of 0.78 (95% CI 0.77-0.79), positive predictive value (PPV) of 0.12 (95% CI 0.10-0.14), negative predictive value (NPV) of 0.99 (95% CI 0.99-0.99), a positive likelihood ratio (PLR) of 3.56 (95% CI 3.30-3.85) and a negative likelihood ratio of 0.26 (95% CI 0.21-0.34) in an adult population with undifferentiated medical complaints. Area under the receiver operating characteristic curve (AUC ROC) is 0.86 for the SEPSIS screening tool.

10.3 Strengths of this study

This study is the first of its nature to be conducted in the UK. Several aspects set this work apart from existing international findings. First, the diagnostic 'gold standard' was not simply extracted from hospital records. The classification of high risk of sepsis, as per the 2016 National Institute for Health and Care Excellence (NICE) sepsis guideline,[21] was confirmed utilising clinical data from the ED clinical record. This is significant as it is well documented that sepsis is under reported.[10 11 213] Indeed, diagnoses of sepsis appear under reported in the data employed in this study - 2.0% of patients received an ED diagnosis of sepsis, whereas 10.0% met the 2013 diagnostic criteria for sepsis[15] while 18.5% were classified as "at risk" per the NICE sepsis guideline (6.5% low risk, 8.3% moderate risk and 3.7% high risk).[21]

Second, the classification of high risk of sepsis, as per the NICE sepsis guideline,[21] is a pragmatic, rather than concrete, clinical diagnosis. Patients classified as high risk for sepsis should receive antibiotic therapy and be reviewed by a senior clinician without delay. Treatment for sepsis is initiated before the diagnosis is confirmed. The screening tool developed in this study is similarly intended to identify those patients who should receive early antibiotic therapy due to their high risk of sepsis, rather than attempting to confirm a sepsis diagnosis.

Third, multiple imputation is the *defacto* standard for the management of missing data. Previous studies to develop prehospital sepsis screening tools have failed to manage missing data by multiple imputation. Missing data have been substituted with 'normal' values,[54 108] or a single value was imputed.[107] In this study missing data were imputed prior to development of the SEPSIS screening tool.

Finally, this study employed a large dataset which guards against overfitting, and has established a baseline rate for high risk sepsis in UK ambulance services.

10.4 Limitations of this study

It was not possible to obtain all the clinical variables requested. Because some clinical variables were missing, it was not possible to calculate an accurate Sequential Organ Failure Assessment (SOFA) score. The SOFA score forms the basis of the most recent international consensus definitions for sepsis.[4] It was hoped that performance of the SEPSIS screening tool could be evaluated against the SOFA score in addition to the NICE high risk of sepsis category to determine how well it performs as a diagnostic aide.

Sadly, there were several prolonged delays obtaining the data to develop the SEPSIS screening tool. Consequently, there was insufficient time to conduct a clinical validation study to assess performance of the SEPSIS screening tool in clinical practice by ambulance clinicians. It remains to be seen how well the SEPSIS screening tool performs in clinical practice.

Finally, the data used in this study derives from one geographic area, consequently performance of the SEPSIS screening tool may be different in other areas of the UK. That is, the SEPSIS screening tool may lack generalisability.

10.5 Implications for practice/policy

If ambulance services are to make a positive contribution to the management of sepsis patients it is vital that ambulance service clinicians receive appropriate education and training to enable them to identify septic patients, and that Emergency Departments act upon pre-alert information passed by ambulance clinicians. Optimal use of National Health Service (NHS) resources is dependent upon accurate diagnosis by ambulance clinicians; over-triage by ambulance clinicians will place unnecessary burden on limited high acuity Emergency

Department resources, whereas under-triage risks acutely ill patients 'slipping through the net', delaying time to diagnosis and potentially worsening outcomes for these patients.

It would be beneficial to establish a mechanism to enable clinical feedback from hospital services to ambulance services. At present NHS Ambulance Trusts design the provision of services in co-operation with partner agencies, but seldom receive feedback from hospitals regarding how well ambulance clinicians are performing within these pathways. That is, data are often available regarding the numbers of patients managed within a pathway (as the ambulance trust can track these data themselves from within their own clinical records). However, there is little data concerning the appropriate use of the pathway. For example, in the care of stroke patients, ambulance crews identify FAST-positive patients and transport them to hyper acute stroke centres. Ambulance services can measure their performance with respect to the numbers of patients managed using these pathways, documentation standards for patients who are fast positive and time intervals for patients transported under these pathways. However, ambulance crews are not informed of the hospital diagnosis. Ambulance clinicians receive minimal feedback concerning diagnostic accuracy. Establishing formal data sharing mechanisms between acute Trusts and the Ambulance Services has the potential to improve care provided by ambulance services over a range of conditions, not just for sepsis patients.

Ongoing monitoring, reported at a national level, designed to capture the management of sepsis patients by NHS Ambulance Services. For example, development of an Ambulance Service Clinical Performance Indicator for Sepsis will be essential to ensure Ambulance Services focus sufficient resources to proactively manage this population.

10.6 Implications for further research

Further research will be required to validate the SEPSIS screening tool for use by ambulance clinicians. It is essential to ensure that ambulance clinicians can reliably identify sepsis patients, before considering the introduction of new treatments such as intravenous antimicrobial therapy.

There is a paucity of evidence addressing sepsis in UK Ambulance Services. It is essential to establish baseline data to determine the incidence and severity of sepsis patients managed by NHS Ambulance Services. This knowledge will help inform the design of education packages, the potential need for investment in equipment such as point of care monitoring, the need to introduce new treatments such as intravenous antibiotic therapy and the design of pathways for sepsis patients.

Other potential research topics include the impact of pre-alert strategies and sepsis pathways on outcomes. The evaluation of point of care monitoring equipment, for example; would the introduction of point of care lactate monitoring improve risk stratification and lead to improved patient outcomes? Alternately, is there any benefit from interventions to optimise perfusion, such a fluid resuscitation and inotropic support? Finally, in relation to prehospital antibiotic therapy, are ambulance staff able to obtain sterile blood samples in the prehospital environment? Does prehospital antibiotic therapy lead to improved patient outcomes?

10.7 Conclusion

This thesis represents a four-year journey of learning and discovery. At the outset, I hoped to develop a screening tool that reliably identified sepsis in the prehospital environment. At the conclusion, I realise that identification of sepsis is far more challenging than I first anticipated. In between, I have learnt a great deal of statistics, developed as an academic writer and have learnt to deal with uncertainty. I have evolved. Considerably.

This thesis extends the body of evidence concerning sepsis in the prehospital environment. The systematic review chapter describes the burden of sepsis on ambulance services. It highlights that diagnosis of sepsis, without reference to diagnostic tests available to in-hospital practitioners, is extremely challenging. Several prehospital sepsis screening tools are available, none of which is ideal.

The main body of the thesis describes the derivation and validation of the SEPSIS score, derived from UK ambulance service data. The methods employed are robust and methodologically sound. The resulting SEPSIS score is a useful

addition to the armament of NHS ambulance clinicians. As the SEPSIS score rises, so too does the probability of sepsis. Utilising a low SEPSIS score threshold ensures high sensitivity, and minimises the possibility of 'missing' a patient with sepsis. Utilising a high SEPSIS score threshold ensures high specificity, and minimises the possibility of initiating inappropriate treatments, such as antibiotic therapy.

In addition to describing the development of sepsis screening, this thesis identifies the burden of sepsis on a UK ambulance service, and compares the performance of several potential screening tools. This information is likely to prove useful to commissioners and those with responsibility to ensure that appropriate services are provided to meet patient's needs.

Appendix 1 Published papers

1.1 Identification of adults with sepsis in the prehospital environment: a systematic review

Open Access Research

BMJ Open

Identification of adults with sepsis in the prehospital environment: a systematic review

Michael A Smyth,[1,2,3] Samantha J Brace-McDonnell,[1,4] Gavin D Perkins[1,4]

To cite: Smyth MA, Brace-McDonnell SJ, Perkins GD. Identification of adults with sepsis in the prehospital environment: a systematic review. *BMJ Open* 2016;**6**: e011218. doi:10.1136/bmjopen-2016-011218

► Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2016-011218).

Received 25 January 2016
Revised 17 April 2016
Accepted 26 May 2016

[1]Clinical Trials Unit, University of Warwick, Coventry, UK
[2]Midlands Air Ambulance, Stourbridge, UK
[3]West Midlands Ambulance Service NHS Foundation Trust, Dudley, UK
[4]Heart of England Hospital NHS Foundation Trust, Birmingham, UK

Correspondence to
Michael A Smyth; m.a.smyth@warwick.ac.uk

ABSTRACT

Objective: Early identification of sepsis could enable prompt delivery of key interventions such as fluid resuscitation and antibiotic administration which, in turn, may lead to improved patient outcomes. Limited data indicate that recognition of sepsis by paramedics is often poor. We systematically reviewed the literature on prehospital sepsis screening tools to determine whether they improved sepsis recognition.

Design: Systematic review. The electronic databases MEDLINE, EMBASE, CINAHL, the Cochrane Library and PubMed were systematically searched up to June 2015. In addition, subject experts were contacted.

Setting: Prehospital/emergency medical services (EMS).

Study selection: All studies addressing identification of sepsis (including severe sepsis and septic shock) among adult patients managed by EMS.

Outcome measures: Recognition of sepsis by EMS clinicians.

Results: Owing to considerable variation in the methodological approach adopted and outcome measures reported, a narrative approach to data synthesis was adopted. Three studies addressed development of prehospital sepsis screening tools. Six studies addressed paramedic diagnosis of sepsis with or without use of a prehospital sepsis screening tool.

Conclusions: Recognition of sepsis by ambulance clinicians is poor. The use of screening tools, based on the Surviving Sepsis Campaign diagnostic criteria, improves prehospital sepsis recognition. Screening tools derived from EMS data have been developed, but they have not yet been validated in clinical practice. There is a need to undertake validation studies to determine whether prehospital sepsis screening tools confer any clinical benefit.

Strengths and limitations of this study

- Despite using very broad search criteria, little robust evidence regarding prehospital sepsis screening was identified.
- The studies found employed disparate methodologies, exhibit significant heterogeneity, generally involve small numbers of patients (limiting the precision of reported results) and were invariably of very low quality.
- The conclusions that can be drawn from this systematic review are, therefore, limited and findings should be interpreted with caution.

INTRODUCTION

Sepsis is a systemic response to infection, which may progress to severe sepsis and septic shock.[1] In the UK, there are an estimated 102 000 cases of severe sepsis each year resulting in >37 000 deaths.[2] It has been reported that more than two-thirds of severe sepsis cases are initially seen in the emergency department (ED)[3] and around half of ED sepsis patients arrive by ambulance.[4–9] Patients with sepsis arriving at the ED via emergency medical services (EMS) are likely to be sicker than those arriving by other means,[5 7–10] and up to 80% of patients with severe sepsis admitted to intensive care from the ED will have been transported by EMS.[11]

Although the burden of sepsis upon ambulance services is not well understood, data from Guerra *et al*[12] suggested that 6.9% of EMS transports were for patients with infection. It is further estimated that 8–10% of EMS patients who have infection will be diagnosed with sepsis.[12 13] Following a 10-year observational study, Seymour *et al*[14] reported the incidence of severe sepsis in a North American EMS system to be 3.3 per 100 ambulance transports. Extrapolation of data reported by McClelland and Jones[15] suggests a lower incidence of sepsis cases in one region of the UK, of ~1.8% of EMS calls.

In-hospital data indicate that early identification and initiation of treatment of severe sepsis is associated with reduced mortality.[2 16] It has been argued that there is an opportunity for ambulance clinicians to improve outcomes for this population in the same manner as they do with other time critical, life-threatening conditions such as acute myocardial infarction,[17] stroke[18] and major trauma.[19] Early recognition of sepsis by

ambulance clinicians could reduce time to delivery of a limited number of interventions prior to arrival at the ED; however, evidence suggests that recognition of sepsis by paramedics is often poor.[7 12 20–22] Use of a prehospital sepsis screening tool has been advocated, suggesting that it would lead to improved recognition, and potentially earlier initiation of key interventions such as fluid resuscitation and antibiotic administration prior to arriving at hospital.[5 23 24]

OBJECTIVE

The objective of this study was to determine whether, among adult patients presenting to EMS, the use of a prehospital sepsis screening tool by ambulance clinicians, compared to ambulance clinician judgement alone, improves identification of sepsis.

DESIGN

We followed the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) Working Group methodology[25] to conduct the review and Preferred Reporting Items for Systematic Reviews and Meta-Analyses recommendations to report our findings.[26] The review is registered with the International Prospective Register of Systematic Reviews (CRD42014007654).

SETTING

Adult patients managed by EMS in the prehospital environment.

STUDY SELECTION

Electronic searches

We searched MEDLINE, EMBASE, CINAHL, the Cochrane Library and PubMed. No language restrictions were placed. Conference proceedings/meeting abstracts were included to capture grey literature.

Search terms/search strategy

Search strategies were developed for each database, starting with MEDLINE (see online supplementary appendix 1). The MEDLINE search strategy was adapted for each subsequent database. Initial searches were conducted in July 2014 with a second search completed at the end of June 2015 (including articles published up to 28 June 2015).

Inclusion criteria

- *Language*: no restrictions were placed.
- *Publication type*: original research published in peer-reviewed journals and conference proceedings.
- *Study design*: systematic reviews, meta-analyses, randomised controlled trials, case–control studies, cohort studies and cross-sectional studies.
- *Study population*: adult patients managed by EMS. Populations could comprise a mix of adult and child participants if results were reported separately.
- *Case definition*: no restrictions as to severity of sepsis.

Exclusion criteria

- *Publication type*: narrative/literature reviews, letters, editorials, commentaries, books and book chapters, lectures and addresses, and consensus statements.
- *Study design*: case reports and qualitative studies.
- *Study population*: In-hospital studies. Mixed adult and child population without distinct reporting, child population and animals.

Other

Reference lists of included manuscripts were scrutinised. Subject experts were contacted to identify studies missed by electronic searches.

Data collection and analysis

Studies were screened in two stages. In the first stage, two reviewers (MAS and SJB-M) independently reviewed each citation and abstract against the inclusion criteria. Citations rated as 'include' by either reviewer were considered relevant, and citations rated as 'exclude' by both reviewers were rejected. In the second stage, the full manuscripts of included citations were again independently screened by two reviewers (MAS and SJB-M) rating each manuscript as 'include', 'maybe' or 'exclude' against the inclusion criteria. If both reviewers rated a manuscript as 'include', it was automatically included for critical appraisal. If both reviewers rated a manuscript as 'exclude', it was automatically rejected. Where a manuscript was selected as 'include' by a single reviewer, or was selected as 'maybe' by one or both reviewers, the reviewers discussed if the manuscript should be included or excluded. If the reviewers were unable to agree, a third independent reviewer (GDP) was available to adjudicate.

OUTCOME MEASURES

We included any study that reported prehospital sepsis screening or development of prehospital sepsis screening tools and compared accuracy of prehospital diagnosis with in-hospital diagnosis.

RESULTS

Database searches yielded 4366 citations. Duplicate citations were removed manually within EndNote (V.X7 Thompson Scientific, Carlsbad, California, USA) by a single reviewer (MAS) providing 2958 unique citations. After the first stage of screening, 78 citations were retained and 2880 citations were rejected. Inter-rater agreement for first-stage screening, calculated using Cohen's κ statistic, was 0.87 (95% CI 0.81 to 0.92). During the second stage of screening, 78 manuscripts were reviewed, 70 were discarded following assessment and 8 were retained for critical appraisal. Inter-rater agreement for second-stage screening, calculated using Cohen's κ statistic, was 0.82 (95% CI 0.68 to 0.97).

No additional citations were identified by scrutinising the reference lists of included manuscripts. One additional

study,[15] a manuscript pending publication, was identified by contacting subject experts. In total, nine studies are included in the final analysis (see figure 1 and online supplementary material).

Characteristics of included studies

No randomised controlled trials were identified; all included studies were observational in nature. Three studies were published in abstract form only.[20 27 28] Studies originate from five countries, comprising a total of 147 320 patients. All studies were published in English. The median year of publication was 2013. The data from included studies were extracted and entered into relevant tables by a single reviewer (MAS) and verified by a second reviewer (SJB-M).

Three studies were concerned with derivation of screening tools.[29–31] Six studies addressed the identification of sepsis within EMS.[12 15 20 21 27 28] Collectively, six prehospital screening tools were identified in the course of the review (critical illness score,[32] Prehospital Recognition of Severe Sepsis (PRESS) score,[31] Prehospital Early Sepsis Detection (PRESEP) score,[30] Robson tool,[21 30] modified Robson tool[15] and BAS 90-30-90[21 30]); a single study reported the accuracy of the Modified Early Warning Score (MEWS).[30] None of the studies were prospective and no studies were designed specifically to validate a prehospital sepsis screening tool in clinical practice.

All studies used hospital sepsis diagnosis as the reference standard; however, hospital diagnosis was variably determined by Surviving Sepsis Campaign diagnostic criteria, International Classification of Disease coding, ED diagnosis (without description of how diagnosis was determined) or discharge diagnosis (without description of how diagnosis was determined).

Risk of bias

Bias within observational studies was assessed across the domains of failure to develop and apply appropriate eligibility criteria (inclusion of control population), flawed measurement of exposure and outcome, failure to adequately control confounding and incomplete

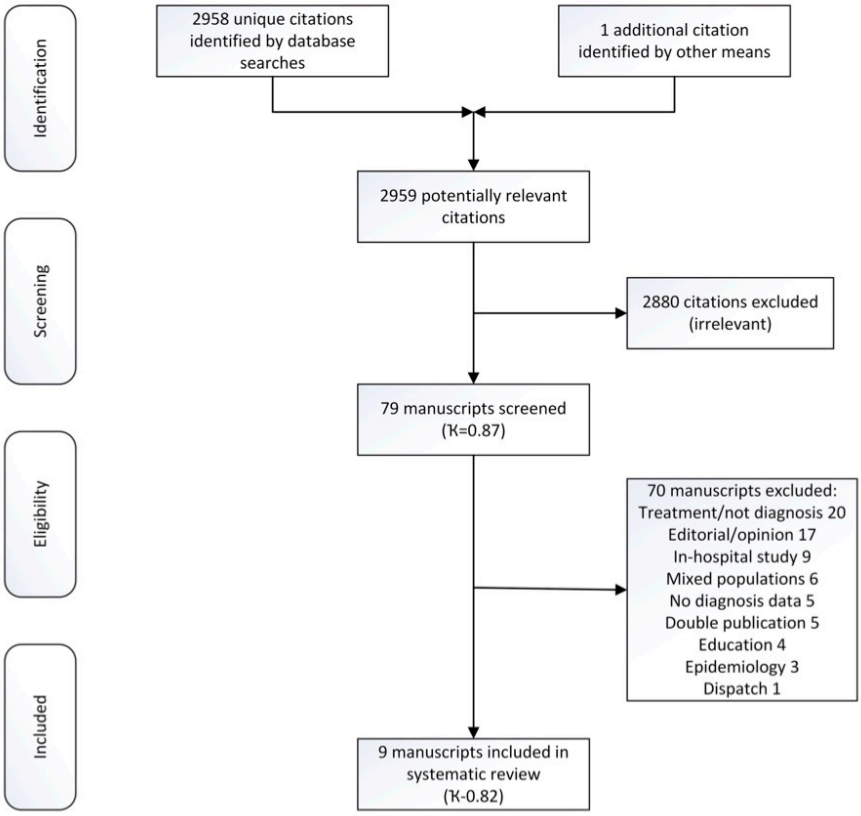


Figure 1 PRISMA flow chart.

follow-up. Two reviewers (MAS and SJB-M) independently assessed each article across the bias domains with each being rated as high risk, low risk or level of risk unclear as per GRADE recommendations.[33] Studies with high risk in one or more domains were considered to be at high risk of bias overall. Similarly studies with low risk for all domains were considered to be at low risk of bias overall. Otherwise, studies were considered to have an unclear risk of bias. Risk of bias assessments are reported in table 1.

Quality of evidence

Study design informed initial quality assumptions. No randomised controlled trials were identified. Non-randomised (observational) studies were initially presumed to be 'low quality'. Two reviewers (MAS and SJB-M) appraised each study across the five core GRADE domains of risk of bias,[33] inconsistency,[34] indirectness,[35] imprecision[36] and other considerations (including publication bias)[37] (see online supplementary material). Where concerns were identified, it lowered the overall quality assumptions. Similarly, quality could have been adjusted upward if, for example, a large treatment effect or dose–response had been noted, which subsequently raised our confidence in the estimate of effect.[38] Quality of evidence, across each outcome of interest, is reported as follows (table 2):

- *High quality*: We are very confident that the true effect lies close to that of the estimate of effect.
- *Moderate quality*: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of effect, but there is a possibility that it is substantially different.
- *Low quality*: Our confidence in the effect is limited: the true effect may be substantially different from the estimate of the effect.
- *Very low quality*: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Data synthesis

There was considerable variation in the methodological approach adopted across the studies as well as the outcome measures reported. The majority of studies identified involve limited numbers of participants, without control and intervention cohorts. Because of these differences, the studies did not answer a unique research question; thus, meta-analysis was not appropriate. A narrative approach to data synthesis was adopted.

Derivation of prehospital sepsis screening tools

We identified very low-quality evidence (downgraded for indirectness and imprecision), from three observational studies,[30–32] addressing derivation of prehospital sepsis screening tools (see table 2). Each of the studies adopted a similar approach to screening tool development. Identification of candidate predictors varied

Table 1 Risk of bias

Author (year)	Industry funding	Eligibility criteria	Exposure/outcome	Confounding	Follow-up
Seymour *et al* (2010)	No	+	+	+	+
Polito *et al* (2015)	No	+	+	+	+
Bayer *et al* (2015)	No	+	+	+	+
Erwin *et al* (2011)	No	-	-	-	?
Shuih *et al* (2012)	No	?	-	?	+
Guerra *et al* (2013)	No	+	+	-	-
Travers *et al* (2013)	No	?	?	+	+
Wallgren *et al* (2014)	No	+	-	-	+
McClelland and Jones (2015)	No	?	?	-	?

-, high risk; +, low risk; ?, risk unclear.

Table 2 Summary of findings

No. of studies	No. of patients	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Findings	Level of evidence
Derivation of prehospital sepsis screening tools									
3	145, 843	Non-RCT	None	None	Not serious*	Very serious†	None	Seymour *et al*[29 32] CIS: risk of sepsis 0.76 (95% CI 0.75 to 0.77) Polito *et al*[31] PRESS score: sensitivity 0.85, specificity 0.47, PPV 0.19, NPV 0.96 (95% CI not reported). PRESS score ≥3; sensitivity 0.81; specificity 0.63 Bayer *et al*[60] PRESEP score: sensitivity 0.85 (95% CI 0.77 to 0.92), specificity 0.86 (95% CI 0.82 to 0.90), PPV 0.66, NPV 0.95	⊕⊙⊙⊙ very low
Sepsis recognition by EMS (using a screening tool)									
2	161	Non-RCT	Very serious‡	None	Not serious§	Very serious¶	Very serious**	Guerra *et al*[12] Screening based on SSC criteria identified 32/67 patients with sepsis (47.8%) (95% CI not reported) McClelland and Jones[15] Screening using modified Robson tool. Sensitivity and specificity for sepsis 43% (95% CI 28% to 58%) and 14% (95% CI 0% to 40%), respectively. Sensitivity and specificity for severe sepsis 30% (95% CI 12% to 47%) and 77% (95% CI 60% to 95%)	⊕⊙⊙⊙ very low
Retrospective application of EMS data to screening tool by researcher									
2	728	Non-RCT	Very serious‡	None	Not serious§	Very serious¶	None	Wallgren *et al*[21] Retrospective application of two different screening tools in comparison to clinical judgement. For sepsis, Robson tool: sensitivity 75% (p<0.001), BAS 90-30-90: sensitivity 43% (p<0.001), clinical judgement: 12% accuracy (95% CI not reported). For severe sepsis, Robson tool: sensitivity 93% (p<0.001), BAS 90-30-90: sensitivity 70% (p<0.001), clinical judgement: 17% accuracy (95% CI not reported) Bayer *et al*[60] Retrospective application of three different screening tools. (Modified) Robson tool: sensitivity 0.95, specificity 0.43, PPV 0.32, NPV 0.97. BAS 90-30-90: sensitivity 0.62, specificity 0.83, PPV 0.51, NPV 0.89. MEWS ≥4 sensitivity 0.74, specificity 0.75, PPV 0.45, NPV 0.91 (95% CI not reported)	⊕⊙⊙⊙ very low

Continued

Table 2 Continued

No. of studies	No. of patients	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Findings	Level of evidence
Sepsis recognition by EMS (use of screening tool not reported)									
3	963	Non-RCT	Very serious‡	None	Not serious§	Very serious¶	Very serious††	Erwin *et al*[20] Screening based on SSC criteria. For sepsis: sensitivity 33% (95% CI 18% to 53%), specificity 89% (95% CI 08% to 94%), PPV 50% (95% CI 28% to 72%), NPV 80% (95% CI 70% to 87%). For severe sepsis: sensitivity 20% (95% CI 5% to 51%), specificity 94% (95% CI 87% to 97%), PPV 29% (95% CI 08% to 64%), NPV 91 (95% CI 83% to 95%) Shiuh *et al*[27] Screening based on SSC criteria, also stratified by lactate, lactate ≤4 'sepsis advisory' while lactate ≥4 'sepsis alert'. 74.2% of 'Sepsis Advisory' patients and 76.7% of 'sepsis alert' patients received a hospital diagnosis of severe infection or sepsis (95% CI not reported) Travers *et al*[28] Screening criteria not defined. Specificity 78.85% (95% CI 75.23 to 82.17), sensitivity 73.4% (95% CI 61.40 to 83.05), PPV 30.59% (95% CI 23.76 to 38.11), NPV 95.86% (95% CI 93.61 to 97.49), accuracy 78% (52 true positives, 440 true negatives)	⊕⊙⊙⊙ very low

*Seymour *et al* CIS not specific to sepsis (CIS intended to identify all cases of critical illness). Polito *et al* and Bayer *et al* studies limited to single EMS systems, Bayer *et al* physician-based EMS.
†Polito *et al* failed to report CIs, small sample size in Bayer *et al* study.
‡All studies patient selection/eligibility criteria, exposure/outcome reporting, confounding.
§Guerra *et al*, Erwin *et al* and Shiuh *et al* include lactate measurement (not widely available within EMS). In majority of studies, the population limited to single EMS agency/hospital so limited generalisablity. Bayer *et al* used physician-based EMS.
¶All included studies have small sample sizes, thus imprecise point estimates. In several studies, CIs are not reported.
**Guerra *et al* publication bias likely.
††Published in abstract only, unable to reliably critically appraise.
BAS 90-30-90, systolic blood pressure <90 mm Hg; respiratory rate >30 bpm; SpO_2 <90%; CIS, critical illness score; EMS, Emergency Medical Services; MEWS, Modified Early Warning Score; modified Robson, Robson tool with addition of SpO_2; non-RCT, non-randomised (observational) study; NPV, negative predictive value; PPV, positive predictive value; PRESEP, Prehospital Early Sepsis Detection; SSC, Surviving Sepsis Campaign.

Table 3 Variables used in screening tools

Author (screening tool)	Variable: Respiratory rate*	Heart rate*	Temperature*	LOC†	SpO_2†	Blood pressure†	Lactate†	Blood glucose†	Skin	CBRT	Dispatch category	Location	Age
Seymour (CIS)	•	•			•	•				•			
Polito (PRESS)			•		•	•					•	•	•
Bayer (PRESEP)	•	•	•		•	•							
Wallgren (Robson tool)	•	•	•	•				•					
Wallgren (BAS 90-30-90)	•				•	•							
McClelland (modified Robson tool)	•	•	•	•	•			•					
Bayer (MEWS)	•	•	•	•		•							
Erwin (unnamed)	•	•	•	•			•		•	•			
Guerra (unnamed)	•	•	•			•	•						
Shiuh (unnamed)	•	•	•				•						

*SIRS criteria.
†Organ dysfunction.
CBRT, capillary bed refill time; CIS, critical illness score; LOC, reduced level of consciousness; MEWS, Modified Early Warning Score; PRESEP, Prehospital Early Sepsis Detection; SIRS, systemic inflammatory response syndrome; SpO_2, oxygen saturations.

Table 4 Performance of screening tools

Author	Sensitivity	Specificity	PPV	NPV
Seymour (CIS)	0.76 (95% CI 0.75 to 0.77)	Not reported	Not reported	Not reported
Polito (PRESS)	0.85 (95% CI not reported)	0.47 (95% CI not reported)	0.19 (95% CI not reported)	0.96 (95% CI not reported)
Bayer (PRESEP)	0.85 (95% CI 0.77 to 0.92)	0.86 (95% CI 0.82 to 0.90)	0.63 (95% CI not reported)	0.95 (95% CI not reported)
McClelland (sepsis) (modified Robson tool)	0.43 (95% CI 0.28 to 0.58)	0.14 (95% CI 0 to 0.40)	Not reported	Not reported
McClelland (severe sepsis) (modified Robson tool)	0.30 (95% CI 0.12 to 0.47)	0.77 (95% CI 0.60 to 0.95)	Not reported	Not reported
Bayer (modified Robson tool)	0.95 (95% CI not reported)	0.43 (95% CI not reported)	0.32 (95% CI not reported)	0.97 (95% CI not reported)
Wallgren (sepsis) (Robson tool)	0.75 (95% CI not reported)	Not reported	Not reported	Not reported
Wallgren (severe sepsis) (Robson tool)	0.93 (95% CI not reported)	Not reported	Not reported	Not reported
Bayer (BAS 90-30-90)	0.62 (95% CI not reported)	0.83 (95% CI not reported)	0.51 (95% CI not reported)	0.89 (95% CI not reported)
Wallgren (sepsis) (BAS 90-30-90)	0.73 (95% CI not reported)	Not reported	Not reported	Not reported
Wallgren (severe sepsis) (BAS 90-30-90)	0.70 (95% CI not reported)	Not reported	Not reported	Not reported
Bayer (MEWS)	0.74 (95% CI not reported)	0.75 (95% CI not reported)	0.45 (95% CI not reported)	0.91 (95% CI not reported)
Guerra	0.48 (95% CI not reported)	Not reported	Not reported	Not reported
Erwin (sepsis)	0.33 (95% CI 0.18 to 0.53)	0.89 (95% CI 0.08 to 0.94)	0.50 (95% CI 0.28 to 0.72)	0.80 (95% CI 0.70 to 0.87)
Erwin (severe sepsis)	0.20 (95% CI 0.05 to 0.51)	0.94 (95% CI 0.87 to 0.97)	0.29 (95% CI 0.08 to 0.64)	0.91 (95% CI 0.83 to 0.95)
Shiuh	0.75 (95% CI not reported)	Not reported	Not reported	Not reported
Travers	0.73 (95% CI 0.61 to 0.83)	0.79 (95% CI 0.75 to 0.82)	0.31 (95% CI 0.24 to 0.38)	0.96 (95% CI 0.94 to 0.98)

CIS, critical illness score; MEWS, Modified Early Warning Score; PRESEP, Prehospital Early Sepsis Detection.

slightly between studies; however, once candidate predictors were identified, all studies used univariate logistic regression to determine which candidate predictors were associated with sepsis, followed by multivariable logistic regression, in a stepwise fashion, to build their respective models. Goodness of fit was assessed by Hosmer-Lemeshow test and model performance determined by calculating the area under the receiver operating characteristic curve.[30–32] Variables used in each screening tool are shown in table 3. None of the studies included a validation study of their respective screening tools.

Seymour *et al*[32] developed the critical illness score to predict the risk critical illness among EMS patients. It was not developed to identify sepsis specifically, although the statistical estimates reported in this review relate to sepsis cases only. Their study used the clinical records of 144 913 EMS patients, of whom 4895 had severe sepsis. Polito *et al*[31] derived the PRESS score from a population of 66 439 EMS encounters. The sample studied included 555 patients at risk of sepsis, of whom 75 were noted to have severe sepsis, while Bayer *et al*[30] derived the PRESEP score from a sample of 375 EMS patients, of whom 93 had sepsis (including 60 patients with severe sepsis and 12 patients with septic shock). Accuracy of prehospital sepsis screening tools is presented in table 4.

Sepsis recognition by EMS (using a screening tool)

We identified very low-quality evidence (downgraded for risk of bias, indirectness and imprecision), from two observational studies,[12 15] addressing recognition of sepsis by EMS personnel using a screening tool (see table 2). Guerra *et al*[12] report that emergency medical technicians (EMTs) trained to recognise sepsis correctly identified 32/67 (48%) patients with sepsis, with failure to recognise sepsis in 35/67 (52%) of cases; however, this figure may be misleading. In 5/35 (14%) of cases, the patient's vital signs did not meet the criteria of the sepsis screening tool while in EMS care; in 8/35 (23%) of cases, the patients had cryptic shock but EMTs did not have lactate meters; and in 13/35 (37%) of cases, diagnosis was made by abnormal white cell count (only available in hospital). In 9/35 (26%) of cases, EMTs failed to identify sepsis when sufficient diagnostic criteria were available to them. The high proportion of patients missed due to lack of white cell count highlights a limitation of prehospital sepsis screening tools. Guerra *et al*[12] further reported that among patients with sepsis transported by EMS crews not trained to recognise sepsis, 5/45 (11%) were identified as patients with sepsis.

McClelland and Jones[15] scrutinised the records of all patients with sepsis conveyed by a regional ambulance service to a university hospital to determine whether ambulance clinicians, previously trained in the use of a screening tool, recognised and documented suspected sepsis. The screening tool used was based on the Robson tool amended to include oxygen saturations as an indicator of organ dysfunction. The authors concluded that the use of the screening tool by ambulance clinicians was inconsistent but improved sepsis recognition.

Retrospective application of EMS data to screening tool by researcher

We identified very low-quality evidence (downgraded for risk of bias, indirectness and imprecision), from two observational studies,[21 30] addressing retrospective application of prehospital data to screening tools (see table 2). Wallgren *et al*[21] compared two screening tools (Robson tool and BAS 90-30-90 score) with EMS clinician judgement. The Robson tool performed better than the BAS 90-30-90 score (see table 4). Clinician judgement, defined as 'documentation of suspected sepsis, septicaemia, urosepsis or blood poisoning in the patient's clinical record', was reported to be 11.9% and 16.9% sensitive for sepsis and severe sepsis, respectively. CIs were not reported.

Bayer *et al*[30] compared the performance of their PRESEP score with the MEWS, BAS 90-30-90 and Robson tool reporting that the PRESEP score surpassed MEWS and BAS 90-30-90 for sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The Robson tool showed better sensitivity; however, the PRESEP tool had better specificity. Furthermore, the PRESEP score showed better PPV and comparable NPV than the Robson tool (see table 4).

Sepsis recognition by EMS (use of screening tool not reported)

We identified very low-quality evidence (downgraded for risk of bias, indirectness, imprecision and abstract only publication), from three observational studies,[20 27 28] addressing accuracy of paramedic diagnosis of sepsis in clinical practice (see table 2). All three studies were published in abstract and it is not clear if paramedics used a screening tool or if they received any training to improve sepsis recognition.

Erwin *et al*[20] compared paramedic diagnosis of sepsis and severe sepsis with physician diagnosis (see table 4). The level of agreement between paramedics and physicians was low (κ=0.25 and 0.16, respectively). These results lead the authors to conclude that sepsis criteria were more useful for ruling-out sepsis than diagnosing sepsis.

In the study by Shiuh *et al*,[27] EMS crews stratified patients with sepsis according to prehospital lactate readings. If patients had a lactate >4 mmol/L, paramedic crews provided the hospital with an 'alert' message, whereas if the lactate was in the range of 2.5–3.9 mmol/L, they provided the hospital with an 'inform' message prior to, or on, hospital arrival. They reported data for 219 patients with sepsis for whom a lactate reading was available; they did not report data for those patients where a lactate reading was not known/unavailable (see table 4).

Travers *et al*[28] compared accuracy of paramedic sepsis diagnosis in 629 cases. Thermometry was not available to

paramedics to confirm body temperature with any degree of accuracy. Paramedic diagnosis agreed with physician diagnosis in 78% of cases. This is the largest paramedic diagnostic accuracy study, but unfortunately detail is lacking.

DISCUSSION

The studies identified provide low-quality or very low-quality evidence to suggest that accuracy of prehospital sepsis recognition by ambulance clinicians varies considerably. This variation could have numerous causes. In many areas, paramedic education programmes have not focused sufficient attention on sepsis as a clinical syndrome and paramedic knowledge of sepsis is often poor.[5 39–41] It is possible that ambulance clinicians encounter patients with sepsis earlier in their clinical course, before they become seriously ill, and it is also not known if in-hospital and prehospital clinical assessments, such as blood pressure, correlate in patients with sepsis. An additional factor may be that routine in-hospital tests such as white cell count and lactate are not commonly used within EMS, which may limit the ability to extrapolate from in-hospital studies.

The majority of the prehospital sepsis screening tools rely upon the Surviving Sepsis Campaign systemic inflammatory response syndrome (SIRS) criteria which were initially described to improve sepsis recognition in the ED and intensive care environments. Although SIRS describe physiological signs marking the transition from infection to sepsis, they lack specificity for sepsis. SIRS are observable following a wide variety of insults other than infection, leading some to question the value of SIRS to identify sepsis.[42 43] Churpek *et al*[44] recently demonstrated that SIRS criteria were not reliable predictors of sepsis or mortality in the ward setting. Use of SIRS criteria to identify sepsis in the prehospital environment may therefore be equally ineffective.

The three studies documenting the development of prehospital screening tools for sepsis included more organ dysfunction criteria and also included non-SIRS variables (see table 3). Among these, tools sensitivity for severe sepsis ranged from 0.76 to 0.85, while specificity ranged from 0.47 to 0.86; they appear to perform better than tools based on the SIRS diagnostic criteria (see table 4); however, none have been clinically validated.

Although nine studies were identified in the course of this review, only five were concerned with screening of patients in clinical practice by EMS clinicians.[12 15 20 27 28] These studies enrolled a total of 1123 patients, over half of whom (675) were in the Travers *et al*[28] study. Given the very limited number of participants in the remaining studies (range 49–183), it is unlikely that reported point estimates are sufficiently precise to draw conclusions with confidence.

CONCLUSION

The identified studies indicate that sepsis recognition within EMS is highly variable. The majority of screening tools studied in clinical practice favour SIRS criteria which may limit the specificity of these tools. Screening tools derived from EMS data have been developed; these tools appear to include more organ dysfunction variables. Retrospective application of ambulance data to these EMS-derived tools suggests that they may help improve sepsis recognition as they demonstrate similar sensitivity with greater specificity. There is a need to undertake validation studies of EMS-derived sepsis screening tools to determine their efficacy. It remains to be seen if use of a prehospital sepsis screening tool provides any significant clinical benefit.

Twitter Follow Michael Smyth at @quincy516

Contributors MAS developed the protocol, developed and ran the searches, selected and appraised papers, extracted data and drafted the initial manuscript. SJB-M selected and appraised papers, verified extracted data and revised the manuscript for important intellectual content. GDP commented on protocol, searches, evidence appraisal and revised the manuscript for important intellectual content. All authors approved the final manuscript.

Funding This work was supported by a National Institute for Health Research Clinical Doctoral Research Fellowship awarded to MAS (CDRF-2012-05-058).

Competing interests GDP is a NIHR Senior Investigator and Director of Research for the Intensive Care Foundation.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Dellinger RP, Levy MM, Rhodes A, *et al*. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013;41:580–637.
2. Daniels R, Nutbeam T, McNamara G, *et al*. The sepsis six and the severe sepsis resuscitation bundle: a prospective observational cohort study. *Emerg Med J* 2011;28:507–12.
3. Wang HE, Shapiro NI, Angus DC, *et al*. National estimates of severe sepsis in United States emergency departments. *Crit Care Med* 2007;35:1928–36.
4. Studnek JR, Artho MR, Garner CL Jr, *et al*. The impact of emergency medical services on the ED care of severe sepsis. *Am J Emerg Med* 2012;30:51–6.
5. Wang HE, Weaver MD, Shapiro NI, *et al*. Opportunities for emergency medical services care of sepsis. *Resuscitation* 2010;81:193–7.
6. Ibrahim I, Jacobs IG. Can the characteristics of emergency department attendances predict poor hospital outcomes in patients with sepsis? *Singapore Med J* 2013;54:634–8.
7. Groenewoudt M, Roest AA, Leijten FMM, *et al*. Septic patients arriving with emergency medical services: a seriously ill population. *Eur J Emerg Med* 2014;21:330–5.
8. Roest AA, Stoffers J, Pijpers E, *et al*. Ambulance patients with nondocumented sepsis have a high mortality risk: a retrospective study. *Eur J Emerg Med* 2015. doi:10.1097/MEJ.0000000000000302
9. van der Wekken LC, Alam N, Holleman F, *et al*. Epidemiology of sepsis and its recognition by emergency medical services personnel in the Netherlands. *Prehosp Emerg Care* 2016;20:90–6.
10. Femling JK, Weiss SJ, Fullerton L, *et al*. EMS sepsis patients are different from walk-in sepsis patients. *Ann Emerg Med* 2012; 60:S120.
11. Gray A, Ward K, Lees F, *et al*. The epidemiology of adults with severe sepsis and septic shock in Scottish emergency departments. *Emerg Med J* 2013;30:397–401.
12. Guerra WF, Mayfield TR, Meyers MS, *et al*. Early detection and treatment of patients with severe sepsis by prehospital personnel. *J Emerg Med* 2013;44:1116–25.
13. Suffoletto B, Frisch A, Prabhu A, *et al*. Prediction of serious infection during prehospital emergency care. *Prehosp Emerg Care* 2011;15:325–30.

14. Seymour CW, Rea TD, Kahn JM, *et al.* Severe sepsis in pre-hospital emergency care: analysis of incidence, care, and outcome. *Am J Respir Crit Care Med* 2012;186:1264–71.
15. McClelland G, Jones J. A pilot study exploring the accuracy of pre-hospital sepsis recognition in the North East Ambulance Service. *J Paramedic Pract* 2015;7:459–65.
16. Gu WJ, Wang F, Bakker J, *et al.* The effect of goal-directed therapy on mortality in patients with sepsis—earlier is better: a meta-analysis of randomized controlled trials. *Crit Care* 2014;18:570.
17. Moyer P, Ornato JP, Brady WJ Jr, *et al.* Development of systems of care for ST-elevation myocardial infarction patients: the emergency medical services and emergency department perspective. *Circulation* 2007;116:e43–8.
18. Abdullah AR, Smith EE, Biddinger PD, *et al.* Advance hospital notification by EMS in acute stroke is associated with shorter door-to-computed tomography time and increased likelihood of administration of tissue-plasminogen activator. *Prehosp Emerg Care* 2008;12:426–31.
19. American College of Emergency Physicians. Trauma care systems development, evaluation, and funding. Policy statement. *Ann Emerg Med* 2012;60:249–50.
20. Erwin A, Salzman J, Wewerka S, *et al.* Can paramedics accurately diagnose sepsis and severe sepsis in the field? *Acad Emerg Med* 2011;1:S23–4.
21. Wallgren UM, Castrén M, Svensson AE, *et al.* Identification of adult septic patients in the prehospital setting: a comparison of two screening tools and clinical judgment. *Eur J Emerg Med* 2014;21:260–5.
22. Polito C, Isakov A, Yancey A, *et al.* Detection of severe sepsis in the pre-hospital setting: how are we doing? *Crit Care Med* 2014;1:A1603.
23. Boardman S, Richmond C, Robson W, *et al.* Prehospital management of a patient with severe sepsis. *J Paramedic Pract* 2009;1:183–8.
24. Robson W, Nutbeam T, Daniels R. Sepsis: a need for prehospital intervention? *Emerg Med J* 2009;26:535–8.
25. Schünemann H, Brozek J, Guyatt G, *et al.* GRADE handbook. Secondary GRADE handbook, 2013. http://www.guidelinedevelopment.org/handbook/
26. Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006–12.
27. Shiuh T, Sweeney T, Rupp R, *et al.* An emergency medical services sepsis protocol with point-of-care lactate accurately identifies out-of-hospital patients with severe infection and sepsis. *Ann Emerg Med* 2012;1:S44.
28. Travers A, Green R, Cain E, *et al.* Can paramedics diagnose sepsis in the prehospital setting? A feasibility study. *Can J Emerg Med* 2013;15:S55.
29. Seymour CW, Kahn JM, Cooke CR, *et al.* Out-of-hospital prediction of critical illness in a population-based cohort. *Am J Respir Crit Care Med* 2010;181 (Meeting Abstracts).
30. Bayer O, Schwarzkopf D, Stumme C, *et al.* An early warning scoring system to identify septic patients in the prehospital setting: the PRESEP score. *Acad Emerg Med* 2015;22:868–71.
31. Polito CC, Isakov A, Yancey AH II, *et al.* Prehospital recognition of severe sepsis: development and validation of a novel emergency medical services screening tool. *Am J Emerg Med* 2015;33:1119–25.
32. Seymour CW, Kahn JM, Cooke CR, *et al.* Prediction of critical illness during out-of-hospital emergency care. *JAMA* 2010;304:747–54.
33. Guyatt GH, Oxman AD, Vist G, *et al.* GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407–15.
34. Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol* 2011;64:1294–302.
35. Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 2011;64:1303–10.
36. Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 2011;64:1283–93.
37. Guyatt GH, Oxman AD, Montori V, *et al.* GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64:1277–82.
38. Guyatt GH, Oxman AD, Sultan S, *et al.* GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011;64:1311–16.
39. Seymour CW, Carlbom D, Engelberg RA, *et al.* Understanding of sepsis among emergency medical services: a survey study. *J Emerg Med* 2012;42:666–77.
40. Báez AA, Hanudel P, Perez MT, *et al.* Prehospital Sepsis Project (PSP): knowledge and attitudes of United States advanced out-of-hospital care providers. *Prehosp Disaster Med* 2013;28:104–6.
41. Shime N. A survey of the competency of ambulance service personnel in the diagnosis and management of sepsis. *J Emerg Med* 2015;49:147–51.
42. Kaukonen KM, Bailey M, Pilcher D, *et al.* Systemic inflammatory response syndrome criteria in defining severe sepsis. *N Engl J Med* 2015;372:1629–38.
43. Vincent JL, Opal SM, Marshall JC, *et al.* Sepsis definitions: time for change. *Lancet* 2013;381:774–5.
44. Churpek MM, Zadravecz FJ, Winslow C, *et al.* Incidence and prognostic value of the systemic inflammatory response syndrome and organ dysfunctions in ward patients. *Am J Respir Crit Care Med* 2015;192:958–64.

1.2 Impact of prehospital care on outcomes in sepsis: a systematic review

SYSTEMATIC REVIEW

Impact of Prehospital Care on Outcomes in Sepsis: A Systematic Review

Michael A Smyth, MSc
Samantha J Brace-McDonnell, MSc
Gavin D Perkins, MD

University of Warwick, Clinical Trials Unit, Coventry, England

Section Editor: Mark I. Langdorf, MD, MHPE
Submission history: Submitted February 23, 2016; Revision received April 14, 2016; Accepted May 3, 2016
Electronically published July 5, 2016
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2016.5.30172

Introduction: Sepsis is a common and potentially life-threatening response to an infection. International treatment guidelines for sepsis advocate that treatment be initiated at the earliest possible opportunity. It is not yet clear if very early intervention by ambulance clinicians prior to arrival at hospital leads to improved clinical outcomes among sepsis patients.

Methoda: We systematically searched the electronic databases MEDLINE, EMBASE, CINAHL, the Cochrane Library and PubMed up to June 2015. In addition, subject experts were contacted. We adopted the GRADE (grading recommendations assessment, development and evaluation) methodology to conduct the review and follow PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) recommendations to report findings.

Results: Nine studies met the eligibility criteria – one study was a randomized controlled trial while the remaining studies were observational in nature. There was considerable variation in the methodological approaches adopted and outcome measures reported across the studies. Because of these differences, the studies did not answer a unique research question and meta-analysis was not appropriate. A narrative approach to data synthesis was adopted.

Conclusion: There is little robust evidence addressing the impact of prehospital interventions on outcomes in sepsis. That which is available is of low quality and indicates that prehospital interventions have limited impact on outcomes in sepsis beyond improving process outcomes and expediting the patient's passage through the emergency care pathway. Evidence indicating that prehospital antibiotic therapy and fluid resuscitation improve patient outcomes is currently lacking. [West J Emerg Med. 2017;17(4)427-437.]

INTRODUCTION

Sepsis is a common and potentially life-threatening response to an infection.[1] There are an estimated 150,000 cases of severe sepsis resulting in more than 44,000 deaths each year in the United Kingdom (UK).[2] It has been reported that over 70% of sepsis cases stem from the community[3] with one study suggesting two-thirds of severe sepsis cases are initially seen in the emergency department (ED).[2] Approximately half of all ED sepsis patients will arrive via emergency medical services (EMS).[5-10] Sepsis patients transported to the ED by EMS are likely to be sicker than those arriving by other means,[6,8-11] with up to 80% of severe sepsis patients admitted to intensive care from the ED having been transported by EMS.[7,12]

International treatment guidelines for sepsis advocate that treatment be initiated at the earliest possible opportunity.[1] It has been argued that early intervention by ambulance clinicians prior to arrival at the ED may lead to improved outcomes among sepsis patients[13] in the same manner as EMS intervention has helped to improve outcomes for other time critical, life-threatening conditions such as acute myocardial

infarction[14], stroke[15], and major trauma.[16]

METHODS

This systematic review addresses the impact of prehospital care on outcomes among patients with sepsis. The review adopted the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) methodology[17] and is reported consistent with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations.[18]

Inclusion Criteria

Studies were eligible for inclusion if they reported the impact of prehospital care among adult patients with suspected sepsis (including severe sepsis and septic shock). Outcomes of interest include time to early goal-directed therapy (EGDT) related targets, admission to intensive care unit (ICU), length of stay and mortality. We included conference proceedings/ meeting abstracts to capture gray literature.

Search Strategy

Electronic Searches

We systematically searched MEDLINE, EMBASE, CINAHL, the Cochrane Library and PubMed. No language restrictions were employed.

Search Terms/Search Strategy

Search strategies were based upon the terms below: (Sepsis OR septic OR septic?emia OR systemic adj inflammatory adj response adj syndrome OR SIRS OR septic adj shock OR hypotension adj induced adj hypoperfusion OR cryptic adj shock OR bacterial adj infection) AND (emergency adj medical adj service OR EMS OR HEMS OR emergency adj medical adj technician OR EMT OR paramedic OR pre-hospital OR prehospital OR pre adj hospital OR out-of-hospital OR out adj of adj hospital OR OOH OR Ambulance).

The initial MEDLINE search was conducted in July 2014 and adapted for each subsequent database. The searches were repeated in June 2015 to identify recent publications.

Other

We contacted subject experts and scrutinized reference lists of included manuscripts in order to identify any missed studies.

Data Collection And Analysis

Study Selection

Study selection occurred in two stages. First, two reviewers (MAS and SJBM) independently reviewed each citation and abstract against the inclusion criteria. Citations rated as 'include' by either reviewer were retained; citations rated as 'exclude' by both reviewers were rejected. Second, full manuscripts of retained citations were independently screened by two reviewers (MAS and SJBM) who rated each manuscript as 'include,' 'maybe,' or 'exclude' against the inclusion criteria. If both reviewers rated a manuscript as 'include' it was included for critical appraisal. If both reviewers rated a manuscript as 'exclude' it was automatically rejected. If the two reviewers had differing opinions, the reviewers discussed the manuscript in order to achieve consensus. If the reviewers were unable to agree following discussion, a third independent reviewer (GDP) was available to adjudicate.

Risk Of Bias

For randomized controlled trials, we assessed risk of bias across the following domains: lack of allocation concealment, lack of blinding, incomplete accounting of patients and outcome events, selective outcome reporting bias and other limitations such as stopping a trial early for benefit. For observational studies, bias was assessed across the domains of failure to develop and apply appropriate eligibility criteria (inclusion of control population), flawed measurement of exposure and outcome, failure to adequately control confounding and incomplete follow up.

All papers were assessed across their respective domains with each being categorized as either high risk, low risk or level of risk unclear as per GRADE recommendations.[19] We considered studies categorized as high risk in any domain to be at high risk of bias overall. Studies categorized as low risk across all domains were considered to be at low risk of bias overall. Studies with a combination of low and unclear risk across domains were considered to have an unclear risk of bias overall.

Quality Of Evidence

We determined quality of evidence according to the GRADE framework. Study design informed initial quality presumptions; randomized controlled trials were initially presumed to be 'high quality,' while observational studies (non-randomized studies) were initially presumed to be 'low quality.' Two reviewers (MAS and SJBM) appraised each paper across the five core GRADE domains of risk of bias,[19] inconsistency,[20] indirectness,[21] imprecision[22] and other considerations (including publication bias).[23] If any concerns were identified quality of evidence was adjusted downward. Similarly, quality could be adjusted upward if, for example, a large treatment effect or dose response was noted, that subsequently raised confidence in the estimate of effect.[24] Ultimately each study is rated as follows:

- High quality: We are very confident that the true effect lies close to that of the estimate of effect.
- Moderate quality: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of effect, but there is a possibility that it is substantially different.
- Low quality: Our confidence in the effect is limited: the true effect may be substantially different from the

estimate of the effect.
- Very low quality: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

RESULTS

Study Inclusion

Database searches yielded 4,366 citations. Duplicate citations were removed manually within EndNote® (version X7 Thompson Scientific, Carlsbad, CA) by a single reviewer (MAS) providing 2,958 unique citations. One citation was identified by contacting subject experts. After the first stage of screening 79 citations were retained and 2,880 citations were rejected. Inter-rater agreement for first stage screening, calculated using Cohens kappa statistic, was 0.87 (95% CI [0.81 to 0.92]). During the second stage of screening 79 manuscripts were reviewed; 70 were discarded following assessment and nine were retained for critical appraisal

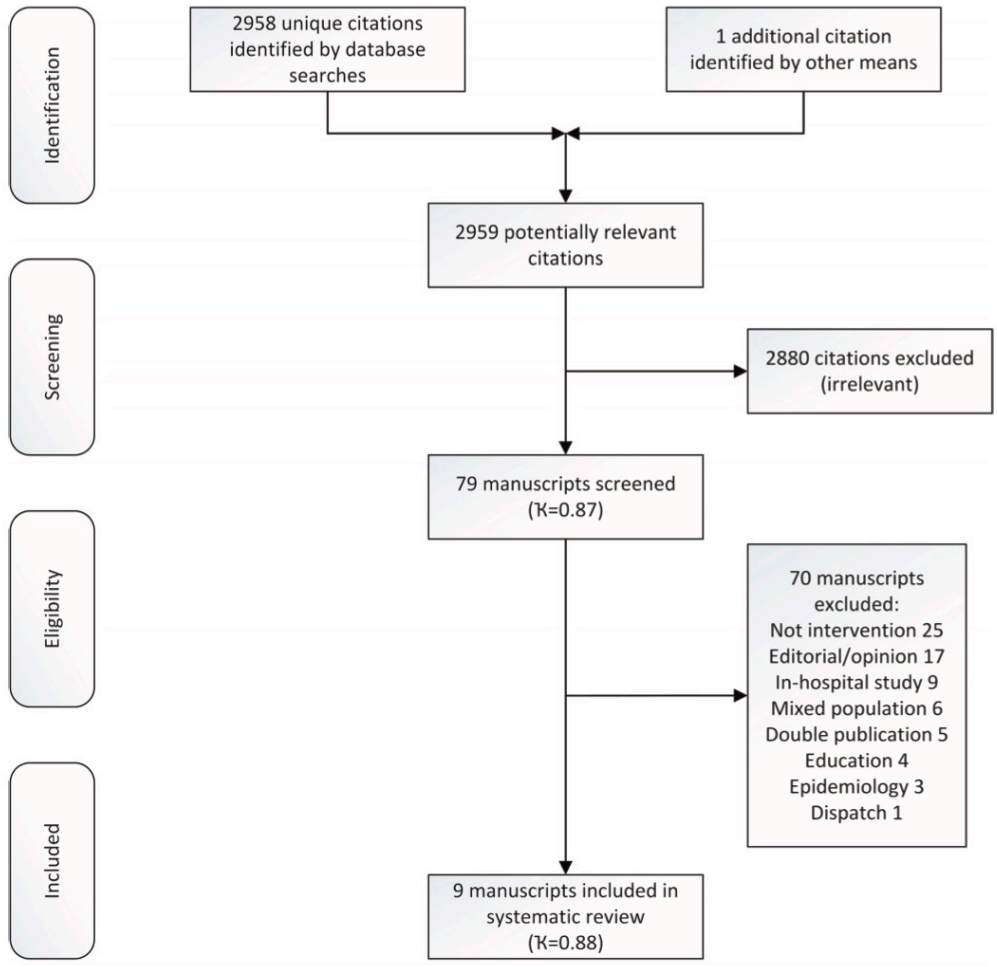


Figure. PRISMA flow chart.
PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Table 1. Characteristics of studies reviewed for quality of evidence regarding whether early intervention by EMS prior to hospital arrival leads to improved clinical outcomes among sepsis patients.

Characteristic	Details
Median year of publication [range]	2013 [2009-2015]
Country of origin [n, (%)]	
Australia	1 (11)
Germany	1 (11)
United Kingdom	1 (11)
United States	6 (67)
Language [n, (%)]	
English	9 (100)
Study design [n, (%)]	
Randomized controlled trials	1 (11)
Non-randomized (observational) studies	8 (89)
Publication type	
Full publication	7 (78)
Abstract publication	2 (22)

EMS, emergency medical services.

(Figure). Inter-rater agreement for second stage screening, calculated using Cohens Kappa, was 0.88 (95% CI [0.72 to 1.0]).

No additional citations were identified by scrutinizing the reference lists of included manuscripts. One additional study,[25] a manuscript pending publication (subsequently published), was identified by contacting subject experts. In total nine studies are included in the final analysis (Figure).

Characteristics Of Included Studies

Characteristics of included studies, comprising 3,470 patients in total, are summarised in the Table.

Risk Of Bias Findings

Risk of bias assessments are reported in Tables 2 and 3.

Quality Of Evidence Findings

We identified very low quality evidence from one randomized controlled trial (downgraded for risk of bias, indirectness and imprecision), and very low quality evidence from eight observational studies (downgraded for risk of bias, indirectness and imprecision across studies, see supplementary information for evidence table with quality assessment.)

Data Synthesis

There was considerable variation in the methodological approach adopted across the studies as well the outcome measures reported. The majority of studies identified involve limited numbers of participants, without comparable control and intervention cohorts. Because of these differences, the studies did not answer a unique research question thus meta-analysis was not appropriate. A narrative approach to data synthesis was adopted.

Data Extraction

The data from included studies were extracted and entered into the evidence table (see Appendix A) and summary of findings table (Table 4) by a single reviewer (MAS) and verified by a second reviewer (SJBM).

ANALYSIS

Antibiotic Therapy

Three studies indicate that ED antibiotic therapy is

Author (year)	Industry funding	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting
Chamberlain (2009)	No	?	?	?	?	+	+

− High risk + Low risk ? Risk unclear

Table 2. Risk of bias (randomized controlled trials).

Author (year)	Industry funding	Eligibility criteria	Exposure/Outcome	Confounding	Follow up
Seymour *et al.* (2010)	no	+	+	?	+
Band *et al.* (2011)	no	+	+	+	+
Studnek *et al.* (2012)	no	+	+	?	+
Bayer *et al.* (2013)	no	?	−	−	+
Guerra *et al.* (2013)	no	+	+	−	−
Femling *et al.* (2014)	no	+	+	?	+
Seymour *et al.* (2014)	no	+	+	?	+
McClelland and Jones (2015)	no	?	?	−	?

− High risk + Low risk ? Risk unclear

Table 3. Risk of bias (non-randomized studies).

administered 30-50 minutes sooner if EMS identify sepsis and inform the receiving clinician of their diagnosis.[5,11,26] However, this finding is not universal – Guerra et al.[27] failed to identify any significant reduction in time to antibiotic therapy (pre-alert: 72.6 minutes Standard Deviation (SD) 59.3 minutes) vs no pre-alert: 98.5 minutes (SD 89.9 minutes), p=0.07). None of the studies concerned with prehospital recognition of sepsis, without concomitant administration of antibiotics, were able to identify any significant improvement in length of stay[11,25,27] or mortality.[11,25-28]

Two studies[29,30] address prehospital administration of antibiotic therapy. Chamberlain[25] reported that antibiotics were delivered 3.4+-2.6 hours sooner while Bayer *et al.*[30] noted that among EMS sepsis patients median time to antibiotics was 19 minutes (IQR 18-24 minutes) from initial emergency call (time of administration was estimated to commence 10 minutes after arriving at scene). Bayer *et al.*[30] do not report interval to hospital nor report time to antibiotics in the ED. Chamberlain[29] suggests that prehospital antibiotic therapy leads to reduced intensive care unit (ICU) stay (Mean ICU stay: 6.8±2.1 days (intervention) vs 11.2±5.2 days (control), p=0.001) and reduced mortality (28-day mortality: 42.4% (intervention) vs 56.7% (control); odds ratio (OR) 0.56; 95% CI [0.32-1.00]). Bayer *et al.*[26] did not report mortality, ICU admission or length-of-stay data.

Intravascular Fluid Therapy

Band *et al.*.[26] reported that arrival by EMS reduces time to initiation of intravascular fluid therapy when compared with those who arrive by privately owned vehicle (POV, EMS: 34 minutes [IQR 10-88 minutes] vs POV: 68 minutes, IQR 25-121 minutes, p≤0.001), but did not improve mortality (adjusted risk ratio [RR] 1.24; 95% CI [0.92-1.66]). Similarly Bayer *et al.*[30] noted that among EMS sepsis patients median time to initiation of Intravenous fluids was 19 minutes (IQR 18-24 minutes) from initial emergency call (time of administration was estimated to commence 10 minutes after arriving at scene), with patients receiving an average of 2.5l intravascular fluid (IQR 1.5–3.01) until admission to the ED. A third study by Guerra *et al.*[27] indicated that early identification of sepsis by EMS was not associated with improved six-hour fluid resuscitation targets in the ED (EMS pre-alert: 42.97 cc/kg (SD 33.23cc/kg) vs no EMS pre-alert: 35.17cc/kg (SD 26.81 cc/kg, p=0.30).

The only study to demonstrate a positive impact following prehospital fluid administration among sepsis patients indicated that prehospital fluids were associated with reduced likelihood of organ failures (adjusted OR 0.58; 95% CI [0.34-0.98]) and reduced hospital mortality (adjusted OR 0.46; 95% CI [0.23-0.88]), but not reduced ICU admission (adjusted OR 0.64; 95% CI [0.37-1.10]).[31] The median volume of prehospital fluid administered in this study was 500mL (IQR 200-1000mL).

Early Goal Directed Therapy (EGDT) Targets

Femling *et al.*[11] reported that patients who arrived at the ED via EMS had shorter time to central line placement (required for central venous pressure monitoring) than those who arrived by other means (EMS: 200 minutes [IQR 89-368 minutes] vs non-EMS: 275 minutes [IQR 122-470 minutes], difference 75 minutes, p<0.01), while Guerra *et al.*[27] noted that when EMS provided a sepsis pre-alert to the hospital the advance notification it did not impact the

№ of studies	№ of patients	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Level of evidence	Findings
Impact of prehospital care upon time to antimicrobial therapy									
1	199	RCT	not serious[1]	none	not serious[2]	very serious[3]	very serious[4]	⊕⊙⊙⊙ very low	[Chamberlain 2009] prehospital antibiotics provided 3.4 ± 2.6 hours sooner (p=0.02).
5	1,927	non-RCT	very serious[5]	none	not serious[2]	very serious[6]	very serious[7]	⊕⊙⊙⊙ very low	[Band 2011] Median time to antibiotics reduced: 116 minutes (IQR 66-199 minutes) EMS vs 152 minutes (IQR 92-252 minutes) 'other means' (p≤0.001). [Studnek 2012] if arriving by EMS vs other means time to antibiotics reduced 111 minutes (EMS) vs 146 minutes (non-EMS); (p=0.001). If EMS recognized and documented sepsis time to antibiotics reduced 70 minutes (documented) vs 122 minutes (not documented) (p=0.003). [Bayer 2013] Median time of administration 19 minutes (IQR 18-24 minutes) after initial emergency call (time of administration estimated as 10 minutes after arriving at scene). [Guerra 2013] No significant reduction in time to antibiotics mean 72.6 minutes (SD 59.3 minutes, pre-alert) vs 98.5 minutes (SD 89.9 minutes, no pre-alert) (p=0.07). [Femling 2014] Time to antibiotics: 87 minutes (EMS, IQR 44-157 minutes) vs 120 minutes (non-EMS, IQR 141-271 minutes), difference 33 minutes (p=0.02).
Impact of prehospital care upon fluid resuscitation									
5	2,697	non-RCT	very serious[5]	none	not serious[2]	very serious[6]	very serious[8]	⊕⊙⊙⊙ very low	[Seymour 2010] patients who received prehospital fluids had shorter time to MAP>65 mm Hg 17/24 (70%, EMS IV fluids) vs 12/26 (44%, no IV fluids), unadjusted RR 1.53 (95% CI [0.9-2.65]), and shorter time to CVP>8 mm H_20 15/25 (60%, EMS IV fluids) vs 17/24 (70%, no IV fluids), unadjusted RR 1.2 (95% CI [0.8-1.8]). [Band 2011] Median time to initiation of IVF reduced: 34 minutes (IQR 10-88) EMS vs 68 minutes (IQR 25-121 minutes) 'other means' of arrival (p≤0.001). [Bayer 2013] Median time of administration 19 minutes (IQR 18-24 minutes) after initial emergency call (time of administration estimated as 10 minutes after arriving at scene). Patients received 2.5L intravascular fluid (IQR 1.5–3.0L) until admitted to the ED. [Guerra 2013] No significant difference in fluid administration by 6 hours 42.97 cc/kg (SD 33.23cc/kg, pre-alert) vs 35.17cc/kg (SD 26.81 cc/kg, no pre-alert, p=0.30). [Seymour 2014] Median prehospital fluid volume 500mL (IQR 200-1000mL).

Table 4. Summary of findings.

RCT, randomized control trial; *EMS*, emergency medical services; *ED*, emergency department; *IQR*, interquartile range; *CI*, confidence interval; *RR*, risk ratio; *MAP*, mean arterial pressure; *CVP*, central venous pressure, *IVF*, intravascular fluid.

Impact of prehospital care upon Early Goal Directed Therapy									
6	2,523	non-RCT	very serious[5]	none	not serious[2]	very serious[6]	very serious[8]	⊕⊙⊙⊙ very low	[Seymour 2010] patients who received prehospital fluids had shorter time to MAP>65 mm Hg 17/24 (70%, EMS IV fluids) vs 12/26 (44%, no IV fluids), unadjusted RR 1.53 (95% CI [0.9-2.65]); shorter time to CVP>8 mm H_20 15/25 (60%, EMS IV fluids) vs 17/24 (70%, no IV fluids), unadjusted RR 1.2 (95% CI [0.8-1.8]); and shorter time to SVC_{O2}>70% 13/24 (54%, EMS IV fluids) vs 9/25 (36%, no IV Fluids), unadjusted RR 1.5 (95% CI [0.8-2.9]). [Studnek 2012] if arriving by EMS vs other means time to EGDT reduced 119 minutes (EMS) vs 160 minutes (non-EMS, p=0.005). If EMS recognised and documented sepsis time to EGDT 69 minutes (documented) vs 131 minutes (not documented, p=0.001). [Guerra 2013] No significant reduction in proportion of patients with central venous line placement 62% (pre-alert) vs 68% (no pre-alert, p=0.54). [Femling 2014] Time to central line: 200 minutes (EMS, IQR 89-368 minutes) vs 275 minutes (non-EMS, IQR 122-470 minutes), difference 75 minutes (p<0.01). [Seymour 2014] Prehospital fluids reduced likelihood of increasing organ failures adjusted OR 0.58 (95% CI [0.34-0.98]). [McClelland 2015] Time to 'sepsis 6': mean 205 minutes (SD 271 minutes, range 10-720 minutes, EMS identified)* vs 120 minutes (SD 110, 17-450 minutes, not identified). (*Includes outlier where the fluid balance chart was not started for 12 hours, excluding this case mean 76 minutes [SD 95 minutes, range 10-240 minutes]).
Impact of prehospital care upon admission									
3	646	non-RCT	very serious[5]	none	not serious[2]	very serious[6]	very serious[8]	⊕⊙⊙⊙ very low	[Guerra 2013] No significant reduction in length of stay: mean 7.3 days (SD 6.8 days, pre-alert) vs 8.4 days (SD 8.8 days, no pre-alert, p=0.65). [Femling 2014] Length of stay: 15 days (IQR 13-17 days, EMS) vs 14 days (IQR 10-17 days, non-EMS), difference 1 day, not significant. [Seymour 2014] Prehospital vascular access reduced ICU admission adjusted OR 0.41 (95% CI [0.24 - 0.70]). [McClelland 2015] ICU admission: 4% (1/23, EMS identified) vs 13% (3/23, not identified).
Impact of prehospital care upon mortality									
5	2,959	non-RCT	very serious[5]	none	not serious[2]	very serious[6]	very serious[8]	⊕⊙⊙⊙ very low	[Band 2011] No significant difference in mortality was noted: adjusted RR 1.24 (95% CI [0.92 - 1.66, p=0.16). [Guerra 2013] If hospital was 'pre-alerted', unadjusted mortality was improved OR 3.19 (95% CI [1.14– 8.88], p=0.04). [Femling 2014] No significant difference in mortality was noted 113/378 (30%, EMS) vs 34/107 (31%, non-EMS), difference 1%, not significant. [Seymour 2014] Prehospital vascular access reduced mortality adjusted OR 0.31 (95% CI [0.17 - 0.57], p<0.01). [McClelland 2015] 3 month mortality 21% (5/24, EMS identified) vs 16% (4/25, not identified).

Table 4. Continued.
RCT, randomized control trial; *EMS*, emergency medical services; *IV,* intravascular; SVC_{O2}, superior vena cava oxygen, *EGDT*, early goal directed therapy; *OR,* odds ratio; *CI,* confidence interval; *ICU,* intensive care unit.

Impact of prehospital antimicrobial therapy on ICU admission									
1	199	RCT	not serious[1]	none	not serious[2]	very serious[3]	very serious[4]	⊕⊙⊙⊙ very low	[Chamberlain 2009] Mean ICU length of stay: reduced 6.8 ± 2.1 days (intervention) vs 11.2 ± 5.2 days (control, p=0.001).
Impact of prehospital antimicrobial therapy on mortality									
1	199	RCT	not serious[1]	none	not serious[2]	very serious[3]	very serious[4]	⊕⊙⊙⊙ very low	[Chamberlain 2009] 28-day mortality reduced: 42.4% (intervention) vs 56.7% (control), OR 0.56 (95% CI [0.32 to 1.00], p=0.049).
Impact of prehospital intravenous fluid therapy on ICU admission									
1	1,350	non-RCT	not serious[9]	none	not serious[2]	none	none	⊕⊙⊙⊙ very low	[Seymour 2014] Prehospital fluids did not reduce likelihood of ICU admission adjusted OR 0.64 (95% CI [0.37-1.10]).
Impact of prehospital intravenous fluid therapy on mortality									
1	1,350	non-RCT	not serious[9]	none	not serious[2]	none	none	⊕⊙⊙⊙ very low	[Seymour 2014] Prehospital fluids reduced hospital mortality adjusted OR 0.46 (95% CI [0.23-0.88], p=0.02).

1. Risk of bias unclear.
2. Single centre study may limit generalizability.
3. Small study numbers limits precision/accuracy.
4. Published in abstract only, insufficient detail to rule out other bias.
5. Concerns relating to eligibility, exposure, confounding, follow-up
6. Small study numbers limits precision/accuracy, failure to report confidence intervals (Guerra)
7. Abstract only publication (Femling), insufficient detail to rule out other bias, Publication bias likely (Guerra)
8. Publication bias likely (Guerra)
9. Risk of bias unclear

Table 4. Continued.

ICU, intensive care unit; *RCT,* non-randomized controlled trial (observational study); *IQR*, interquartile range; *EMS*, emergency medical services; *SD,* standard deviation; *MAP*, mean arterial pressure; *IV,* intravascular; *RR*, risk ratio, *CI,* confidence interval; *CVP*, central venous pressure; *EGDT,* early goal directed therapy; *OR*, odds ratio.

decision to place a central venous catheter (EMS pre-alert: 61% vs no EMS pre-alert: 68%, p=0.54). Although Seymour *et al.*[28] reported that higher proportion of patients achieved a SVC_{O2}>70% within six hours when EMS initiated fluid therapy prior to arriving at the ED, the unadjusted risk ratio found no evidence of a difference (EMS IV fluids: 13/24 (54%) vs no IV fluids: 9/25 (36%), Unadjusted RR 1.5, 95% CI [0.8-2.9]). This same study also identified no improvement in time to MAP>65mmHg (EMS IV fluids: 17/24 (70%) vs no IV fluids: 12/26 (44%), unadjusted RR 1.53 (95% CI [0.9-2.65]), and time to CVP>8 mmH_20 (EMS IV fluids: 15/25 (60%) vs no IV fluids: 17/24 (70%), unadjusted RR 1.2 (95% CI [0.8-1.8]).[28]

Studnek *et al.*[5] reported that if patients arrived by EMS they had shorter times to EGDT than if they arrived by other means (EMS: 119 minutes vs non-EMS: 160 minutes, SD/range not reported, p=0.005). Furthermore, among EMS-transported patients, if EMS documented suspicion of sepsis then time to EGDT was shorter than if they did not document suspicion of sepsis (documented suspicion: 69 minutes vs not documented: 131 minutes, SD/range not reported, p=0.001). McClelland *et al.*[25] similarly reported that time to delivery of the 'Sepsis 6' (administration of supplemental oxygen, intravenous fluids, antibiotics, measurement of venous lactate, urine output, and drawing blood to identify causative pathogen) was shorter if EMS identified sepsis prior to arrival at hospital (EMS identified: mean 205 minutes [SD 271 minutes, range 10-720 minutes] vs not identified: mean 120 minutes [SD 110, 17-450 minutes]). These data points include one outlier where the fluid balance chart was not started for 12 hours. Excluding this case, the mean time to delivery of the 'Sepsis 6' would be 76 minutes (SD 95 minutes, range 10-240 minutes).

DISCUSSION

Very few, if any, EMS systems are capable of delivering the entire initial resuscitation bundle advocated by the Surviving Sepsis Campaign guidelines.[1] Most EMS systems lack the capability to draw blood and analyze the required parameters; in addition some of the technical skills required, such as central line placement, will be beyond the scope of many non-physician providers. It is therefore unreasonable to expect EMS systems to be able to deliver all elements of the initial resuscitation bundle. However, key interventions, such as oxygen therapy, antibiotic administration, fluid resuscitation and measuring venous lactate are possible. Despite the ability of EMS to deliver the aforementioned, recent hospital trials[32-34] have brought into question several of the EGDT objectives. We therefore need to examine carefully the need to extend EMS scope of practice to deliver those elements not routinely practiced, such as measuring venous lactate and administering antibiotics.

Prehospital recognition of sepsis is challenging.[8,27,35] The limited evidence identified suggests the initiation of treatment by EMS may lead to improved process outcomes, i.e. reduces time taken to achieve initial resuscitation targets but is not necessarily associated with improved clinical outcomes.

There is currently no evidence addressing impact of prehospital oxygen therapy in sepsis. The ARISE[33], ProCESS[32] and ProMISe[34] trials have all suggested that the need to rigidly adhere to EGDT may be overstated. Furthermore, a systematic review by Sterling *et al.*[36] indicates that antibiotic administration within the first three hours is not associated with improved patient outcomes.

One study[29] identified during this review suggests that prehospital antibiotics may reduce mortality (OR 0.56 (95% CI [0.32-1.00]), p=0.049); however, this study was published in abstract only and enrolled a limited number of patients (n=198). We cannot therefore be confident that prehospital antibiotics would improve outcomes. The PHANTASi trial (NCT01988428) will hopefully provide further evidence to determine if EMS systems should extend clinical practice to deliver prehospital antibiotic therapy in cases of suspected sepsis.

Fluid therapy is an established clinical practice in many EMS systems. Seymour *et al.*[31] identified that prehospital fluid therapy was associated with both reduced organ failures (OR 0.58, 95% CI [0.34-0.98]) and mortality (OR 0.46, 95% CI [0.23-0.88]); however, the mean volume of fluid administered was only 500ml, considerably below what would normally be administered as part of the initial resuscitation bundle (30mL/kg).[1] This led the authors to question if the reduced mortality was due to the small volume of fluid or indeed if it was associated with process improvements secondary to prehospital recognition of sepsis. The latter argument is strengthened by their finding that placement of an intravenous catheter, without any fluid being administered, was also associated with reduced hospital mortality (OR 0.31, 95% CI [0.17-0.57]).[31]

One further aspect that has not been examined is the influence of EMS system design. Internationally, two distinct EMS systems, the EMT/paramedic (Anglo-American) model and physician (Franco-German) model are observed. Typically physician responders might be expected to have higher clinical acumen than paramedics/EMTs as a result of their longer, more in-depth education and training. In addition they may have greater scope to initiate a broader range of interventions, as well as direct admission to specialist services. These factors could improve recognition and indeed treatment of sepsis before arriving at hospital.

Eight of the included studies were conducted in EMT/paramedic EMS systems[5,11,25-29,31] with a single study, published in abstract only, conducted in a physician-based EMS system.[30] Studies conducted in both system designs suggested reduced times to interventions; however, Bayer *et al.*[30] did not publish data addressing mortality, ICU admission nor length of stay in their EMS physician-based study. Although Bayer *et al.*[30] reported a high proportion of suspected prehospital sepsis cases

were later confirmed in the hospital, they did not report data concerning missed cases making it impossible to determine if EMS physicians are able to accurately identify sepsis patients out of the hospital. Bayer *et al.*[30] did however report a larger mean fluid volume (2.5l intravascular fluid (IQR 1.5–3.0l)),[30] than in the paramedic-based study (mean volume 500mL (IQR 200-1000mL)) reporting this outcome,[31] which may reflect greater understanding of beneficial treatments. With such limited data it is not possible to draw any meaningful conclusions concerning the impact of EMS physicians on outcomes in sepsis.

LIMITATIONS

We employed a broad search strategy in order to capture as much published literature as possible. Inclusion criteria were similarly not restrictive so as to include as much of the evidence base as possible. To the best of our knowledge, this is the first systematic review addressing the impact of prehospital interventions upon outcomes among sepsis patients. Despite using very broad search criteria, little robust evidence regarding the impact of prehospital care of sepsis patients was identified. The studies found employed disparate methodologies, exhibit significant heterogeneity, generally involve small numbers of patients (limiting the precision of reported results) and were invariably of very low quality. The conclusions that can be drawn from this systematic review are therefore limited and findings should be interpreted with caution.

CONCLUSION

There is little robust evidence addressing the impact of prehospital interventions on outcomes in sepsis. That which is available is of very low quality and indicates that prehospital interventions have limited impact on outcomes in sepsis beyond improving process outcomes and expediting the patients passage through the emergency care pathway. Evidence indicating that prehospital antibiotic therapy and fluid resuscitation improve patient outcomes is lacking. Well-conducted studies addressing key clinical interventions, such as antibiotic administration and fluid resuscitation are required.

Address for Correspondence: Michael A Smyth, MSc, University of Warwick, Clinical Trials Unit, University of Warwick, Coventry, CV4 7AL, England. Email: m.a.smyth@warwick.ac.uk.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. Michael A. Smyth and Samantha J. Brace-McDonnell are funded by National Institute for Health Research (NIHR) Clinical Doctoral Research Fellowships. Gavin D Perkins is a NIHR Senior Investigator and Director of Research for the Intensive Care Foundation. The funder played no role in design, analysis, interpretation or reporting of findings. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

REFERENCES

1. Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med.* 2013;41(2):580-637.
2. UK Sepsis Trust. UK Sepsis Trust Homepage. 2015; Available at: http://www.uksepsistrust.org. Accessed Dec 15, 2015.
3. Esteban A, Frutos-Vivar F, Ferguson ND, et al. Sepsis incidence and outcome: contrasting the intensive care unit with the hospital ward. *Crit Care Med.* 2007;35(5):1284-9.
4. Wang HE, Shapiro NI, Angus DC, et al. National estimates of severe sepsis in United States emergency departments. *Crit Care Med.* 2007;35(8):1928-36.
5. Studnek JR, Artho MR, Garner CL, et al. The impact of emergency medical services on the ED care of severe sepsis. *Am J Emerg Med.* 2012;30(1):51-6.
6. Wang HE, Weaver MD, Shapiro NI, et al. Opportunities for Emergency Medical Services care of sepsis. *Resuscitation.* 2010;81(2):193-7.
7. Ibrahim I and Jacobs IG. Can the characteristics of emergency department attendances predict poor hospital outcomes in patients with sepsis? *Singapore Med J.* 2013;54(11):634-8.
8. Groenewoudt M, Roest AA, Leijten FMM, et al. Septic patients arriving with emergency medical services: A seriously ill population. *Eur J Emerg Med.* 2014;21(5):330-5.
9. Roest AA, Stoffers J, Pijpers E, et al. Ambulance patients with nondocumented sepsis have a high mortality risk: a retrospective study. *Eur J Emerg Med.* 2015.
10. van der Wekken LC, Alam N, Holleman F, et al. Epidemiology of Sepsis and Its Recognition by Emergency Medical Services Personnel in the Netherlands. *Prehosp Emerg Care.* 2015.
11. Femling J, Weiss S, Hauswald E, et al. EMS patients and walk-in patients presenting with severe sepsis: differences in management and outcome. *South Med J.* 2014;107(12):751-6.
12. Gray A, Ward K, Lees F, et al. The epidemiology of adults with severe sepsis and septic shock in Scottish emergency departments. *Emerg Med J.* 2013;30(5):397-401.
13. Robson WN, T. and Daniels, R. Sepsis: a need for prehospital intervention? *Emerg Med J.* 2009;26(7):535-8.
14. Moyer P, Ornato JP, Brady WJ, Jr., et al. Development of systems of care for ST-elevation myocardial infarction patients: the emergency medical services and emergency department perspective.

Circulation. 2007;116(2):e43-48.
15. Abdullah AR, Smith EE, Biddinger PD, et al. Advance hospital notification by EMS in acute stroke is associated with shorter door-to-computed tomography time and increased likelihood of administration of tissue-plasminogen activator. *Prehosp Emerg Care*. 2008;12(4):426-31.
16. American College of Emergency Physicians. Trauma care systems development, evaluation, and funding. Policy statement. *Ann Emerg Med*. 2012;60(2):249-50.
17. Schünemann H, Brożek J, Guyatt G, A. O. GRADE Handbook. 2013; Available at: http://www.guidelinedevelopment.org/handbook/. Accessed Nov 21, 2013.
18. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62(10):1006-12.
19. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol*. 2011;64(4):407-15.
20. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. *J Clin Epidemiol*. 2011;64(12):1294-1302.
21. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *J Clin Epidemiol*. 2011;64(12):1303-10.
22. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. *J Clin Epidemiol*. 2011;64(12):1283-93.
23. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. *J Clin Epidemiol*. 2011;64(12):1277-82.
24. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64(12):1311-6.
25. McClelland G and Jones J. A pilot study exploring the accuracy of pre-hospital sepsis recognition in the North East Ambulance Service. *Journal of Paramedic Practice*. 2015;7(9):459-65.
26. Band RA, Gaieski DF, Hylton JH, et al. Arriving by emergency medical services improves time to treatment endpoints for patients with severe sepsis or septic shock. *Acad Emerg Med*. 2011;18(9):934-40.
27. Guerra WF, Mayfield TR, Meyers MS, et al. Early detection and treatment of patients with severe sepsis by prehospital personnel. *J Emerg Med*. 2013;44(6):1116-25.
28. Seymour CW, Cooke CR, Mikkelsen ME, et al. Out-of-hospital fluid in severe sepsis: effect on early resuscitation in the emergency department. *Prehosp Emerg Care*. 2010;14(2):145-52.
29. Chamberlain D. Prehospital administered intravenous antimicrobial protocol for septic shock: A prospective randomized clinical trial. *Critical Care*. 2009;13:S130-S131.
30. Bayer O, Stumme C, Schneider K, et al. Evaluation of 1 year prehospital emergency sepsis care: An observational study. *Infection, Supplement*. 2013;(1):S66.
31. Seymour CW, Cooke CR, Heckbert SR, et al. Prehospital intravenous access and fluid resuscitation in severe sepsis: an observational cohort study. *Crit Care*. 2014;18(5):533.
32. Yealy DM, Kellum JA, Huang DT, et al. A randomized trial of protocol-based care for early septic shock. *N Engl J Med*. 2014;370(18): 1683-93.
33. Peake SL, Delaney A, Bailey M, et al. Goal-directed resuscitation for patients with early septic shock. *N Engl J Med*. 2014;371(16):1496-1506.
34. Mouncey PR, Osborn TM, Power GS, et al. Trial of early, goal-directed resuscitation for septic shock. *N Engl J Med*. 2015;372(14):1301-11.
35. Erwin A, Salzman J, Wewerka S,et al. Can paramedics accurately diagnose sepsis and severe sepsis in the field? *Acad Emerg Med*. 2011;(1):S23-S24.
36. Sterling SA, Miller WR, Pryor J, et al. 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):1907-15.

Appendix 2

2.1 Research ethics committee approval

NHS
Health Research Authority

NRES Committee South Central - Oxford C
Bristol REC Centre
Level 3, Block B
Whitefriars Building
Lewins Mead
Bristol
BS1 2NT
Telephone: 01173 421383

02 April 2014

Mr Michael Smyth
Research Fellow
Warwick Medical School
Gibbet Hill RD
Coventry
CV47AL

Dear Mr Smyth

Study title:	**Development of a prehospital sepsis screening tool for use by ambulance clinicians.**
REC reference:	**14/SC/0163**
IRAS project ID:	**152449**

The Research Ethics Committee reviewed the above application at the meeting held on 28 March 2014. Thank you for attending with Dr Gavin Perkins to discuss the application.

The Committee considered whether a separate tool for recognising sepsis is needed when there are already established clinical criteria for identifying sepsis.

You responded that measurements from in-hospital diagnosis will be compared with those taken out of hospital to see whether these are similar and to look at how accurate the pre-hospital measures are. For example if blood pressure is measured in hospital, the patient and machine are static, but with the movements in an ambulance the measurements are not as precise and the equipment is different, therefore it cannot be assumed that recordings of BP will be the same as when they are taken in a hospital setting.

Furthermore, some of the measures used in hospital such as white blood cell counts cannot be used in a pre-hospital environment, which is why this research will look to build a tool which can be used pre-hospital. For example in a stroke there are well-established pre-hospital tools available such as FAST which focuses on a set of criteria. You hope to develop something similar for sepsis.

The Committee questioned what the 'gold-standard' is; will it be in-hospital diagnosis of sepsis?

You confirmed you will be comparing the outcome predicted by the pre-hospital sepsis screening tool with diagnosis in the emergency department.

NHS
Health Research Authority

The Committee noted that there have been two other similar studies which have developed tools for example NEWS (The National Early Warning Score) and asked what the difference is with this research.

You explained that NEWS predicts risk of death within 24 hours and is not specific to sepsis. The tool developed in this research will be compared with all available tools. You will then aim to look at validation of the tool and a trial once you have developed the best tool to identify sepsis.

The Committee commented that in the Isle of Wight paramedics are already using antibiotics to treat sepsis early, it was questioned how the research fits in with this.

You answered that this is not ongoing active practice in the Isle of Wight; there was a brief trial using a tool based on hospital diagnosis, however you were unsure how successful this was.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details, unless you expressly withhold permission to do so. Publication will be no earlier than three months from the date of this favourable opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to withhold permission to publish, please contact the REC Manager Miss Lauren Allen, nrescommittee.southcentral-oxfordc@nhs.net.

Ethical opinion

The members of the Committee present gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

A Research Ethics Committee established by the Health Research Authority

NHS
Health Research Authority

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database within 6 weeks of recruitment of the first participant (for medical device studies, within the timeline determined by the current registration and publication trees).

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to contest the need for registration they should contact Catherine Blewett (catherineblewett@nhs.net), the HRA does not, however, expect exceptions to be made. Guidance on where to register is provided within IRAS.

Ethical review of research sites

NHS Sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

It is responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Investigator CV	Mike Smyth CV	10 March 2014
Investigator CV	Gavin Perkins CV	12 February 2014
Investigator CV	Matthew Cooke CV	
Other: Data Flow Diagram		10 March 2014
Protocol	1	10 March 2014
REC application		

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the

NHS
Health Research Authority

attached sheet.

There were no declarations of interest.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

Further information is available at National Research Ethics Service website > After Review

14/SC/0163	**Please quote this number on all correspondence**

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

Yours sincerely

pp

Professor Nigel Wellman

A Research Ethics Committee established by the Health Research Authority

NHS
Health Research Authority

Chair

Email: nrescommittee.southcentral-oxfordc@nhs.net

Enclosures: *List of names and professions of members who were present at the meeting and those who submitted written comments*
"After ethical review – guidance for researchers"

Copy to: *Mrs Jane Prewett*
Mrs Gill Price, West Midlands Ambulance Service NHS Foundation Trust
NIGB Ethics & Confidentiality Committee Secretariat

A Research Ethics Committee established by the Health Research Authority

NHS
Health Research Authority

NRES Committee South Central - Oxford C

Attendance at Committee meeting on 28 March 2014

Committee Members:

Name	*Profession*	*Present*	*Notes*
Dr Leonard Brookes	Consultant to the Pharmaceutical Industry	Yes	
Miss Gemma Davison	Solicitor	Yes	
Dr Avinash Gupta	Clinical Research Fellow	Yes	
Mrs Rebekah Howe	Farmer	Yes	
Mrs Vivienne Laurie	Barrister	Yes	
Mrs Susan Lousada	Company Director (Property) & Non-legal member of first-tier tax tribunal	Yes	
Mr Barry Muir	Retired NHS Management Consultant	Yes	
Mrs Rachael Quinn	Nurse Member	Yes	
Professor David Scott	Pharmacist	Yes	
Dr Sabeena Sharma	Consultant Anaesthetist	Yes	
Dr Surjeet Singh	Clinical Trials Coordinator	Yes	
Professor Nigel Wellman	Professor of Health and Human Sciences	Yes	

Also in attendance:

Name	*Position (or reason for attending)*
Miss Lauren Allen	REC Manager
Ms Linda Buckley	Observer
Miss Alice Good	Observer
Miss Deborah King	Observer
Ms Carol Simms	Observer

2.2 Health Research Agency permission to access confidential data

NHS
Health Research Authority
Confidentiality Advisory Group

Mr Michael Smyth
Warwick Clinical Trials Unit
Gibbett Hill Rd
Coventry
CV4 7AL

m.a.smyth@warwick.ac.uk

Skipton House
80 London Road
London
SE1 6LH

Tel: 020 797 22557
Email: HRA.CAG@nhs.net

19 May 2014

Dear Mr Smyth

Study title: **Prehospital Recognition of Sepsis by Ambulance Clinicians (PRoSAiC)**
CAG reference: **CAG 4-03(PR2)2014**
Protocol number: **1.0**
REC number: **14/SC/0163**

Thank you for your research application, submitted for approval under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 to process patient identifiable information without consent. Approved applications enable the data controller to provide specified information to the applicant for the purposes of the relevant activity, without being in breach of the common law duty of confidentiality, although other relevant legislative provisions will still be applicable.

The role of the Confidentiality Advisory Group (CAG) is to review applications submitted under these Regulations and to provide advice to the Health Research Authority on whether an application should be approved, and if so, any relevant conditions. This application was considered via the proportionate review process under criteria 4 – time limited access to undertake record linkage/validation and to pseudonymise the data.

Health Research Authority approval decision

The Health Research Authority, having considered the advice from the Confidentiality Advisory Group as set out below, has determined the following:

1. The application is approved, subject to compliance with the standard and specific conditions of approval.

Context

Purpose of application

This application from the University of Warwick is to produce a 'tool' to allow ambulance staff/paramedics to identify patients with sepsis so that they can be prioritised for treatment on reaching the Emergency Department. The data will be obtained retrospectively only from the West Midlands Ambulance Service NHS Foundation Trust and University Hospital North Staffordshire NHS Foundation Trust, the applying body being Warwick Medical School, Clinical Trials Unit.

A recommendation for class 4 and 6 support is being requested to use patient identifiers to link hospital and ambulance records.

Confidential patient information requested

Access was requested to name, hospital number, date of birth, gender and unit level postcode.

Confidentiality Advisory Group advice

Practicable alternatives

Members considered whether a practicable alternative to the disclosure of patient identifiable data without consent existed, taking into account the cost and technology available in line with Section 251 (4) of the NHS Act 2006.

- Feasibility of consent

Members noted that it would not be feasible to seek consent based on the study design.

- Use of anonymised/pseudonymised data

Members noted that all identifiable data will be removed once the hospital and ambulance records have been linked.

Justification of identifiers

Members noted the identifiers that were requested that the Applicant asserted that those identifiers specified were required to perform the linkage between the hospital and ambulance records.

Fair Processing

Having reviewed the application in terms of the fair processing of data, Members requested that further information be published about the use of patient data within this project, including information about how patients may 'opt-out' of their information being included in this research.

As patients seen within the Accident and Emergency department may never visit the department again it was stated that therefore they would not see the poster specified in the application. It was suggested that an information leaflet be developed and be referenced in each of the organisations websites involved in the study.

It was noted that the West Midlands Ambulance Service NHS Foundation Trust did not include using personal data for research purposes within their current ICO Data Protection registration. The Applicant advised that this would be updated prior to the commencement of any research activity.

The Applicant stated that the fully authorised/signed NIGB Form would be provided with confirmation that no further changes have been made to the form since the submission of this application.

CAG advice conclusion

In line with the considerations above, the CAG agreed that the minimum criteria under the Regulations appeared to have been met and that there was a public interest in research of this nature being conducted, and therefore advised recommending provisional support to the

Health Research Authority, subject to compliance with the specific and standard conditions of support as set out below.

Specific conditions of support

1. Provision of patient information leaflets that reference patients' right to opt-out of their data being included within this project.
2. Receipt of a fully authorised/signed NIGB Form based on the draft form submitted as part of the application with confirmation that there have been no further changes to the NIGB form other than the authorisations of the Declarations.
3. Confirmation that West Midlands Ambulance Service NHS Foundation Trust includes the use of data for research purposes within their ICO Data Protection Registration prior to the start of the research activity.
4. Confirmation of suitable security arrangements via IG Toolkit submission, please see security review requirement section here; http://www.hra.nhs.uk/resources/confidentiality-advisory-group/confidentiality-advisory-group-cag-application-advice/ and contact Exeter.helpdesk@nhs.net with any queries.

Once confirmation of the above conditions has been provided, the response will be reviewed and if satisfactory, the HRA will confirm final approval. Support only comes into effect once this final approval letter has been received.

Please do not hesitate to contact me if you have any queries following this letter. I would be grateful if you could quote the above reference number in all future correspondence.

Reviewed documents

The documents reviewed by Members were:

Document	*Version*	*Date*
IRAS application form		11/03/2014
Protocol	1.0	10/03/2014
Data Flow Diagram		
Caldicott Guardian Support Email		11/04/2014
REC Favourable Opinion Letter		02/04/2014

Membership of the Committee

The members of the Confidentiality Advisory Group who considered this item are listed below.

There were *no* declarations of interest in relation to this item.

With the Group's best wishes for the success of this project.

Yours sincerely

John Robinson
Confidentiality Advisor

Email: HRA.CAG@nhs.net

Enclosures: *List of members who considered application*

Standard conditions of approval

Copy to: *nrescommittee.southcentral-oxfordc@nhs.net*

Confidentiality Advisory Group
Sub Committee Meeting 02/05/2014

Group members

Name	Capacity
Dr Robert Carr	
Dr Patrick Coyle (Chair)	
Dr Murat Soncul	

NHS
Health Research Authority
Confidentiality Advisory Group

Standard conditions of approval

The approval provided by the Health Research Authority is subject to the following standard conditions.

The applicant will ensure that:

1. The specified patient identifiable information is only used for the purpose(s) set out in the application.
2. Confidentiality is preserved and there are no disclosures of information in aggregate or patient level form that may inferentially identify a person, nor will any attempt be made to identify individuals, households or organisations in the data.
3. Requirements of the Statistics and Registration Services Act 2007 are adhered to regarding publication when relevant.
4. All staff with access to patient identifiable information have contractual obligations of confidentiality, enforceable through disciplinary procedures.
5. All staff with access to patient identifiable information have received appropriate ongoing training to ensure they are aware of their responsibilities.
6. Activities are consistent with the Data Protection Act 1998.
7. Audit of data processing by a designated agent is facilitated and supported.
8. The wishes of patients who have withheld or withdrawn their consent are respected.
9. The Confidentiality Advice Team is notified of any significant changes (purpose, data flows, data items, security arrangements) prior to the change occurring.
10. An annual report is provided no later than 12 months from the date of your final confirmation letter.
11. Any breaches of confidentiality / security around this particular flow of data should be reported to CAG within 10 working days, along with remedial actions taken / to be taken.

2.3 Research ethics committee approval (substantial amendment)

NHS
Health Research Authority

NRES Committee South Central - Oxford C
Level 3, Block B
Whitefriars Building
Lewins Mead
Bristol
BS1 2NT
Tel: 01173 421383

06 August 2014

Mr Michael Smyth
Research Fellow
Warwick Medical School
Gibbet Hill RD
Coventry
CV47AL

Dear Mr Smyth

Study title:	**Development of a prehospital sepsis screening tool for use by ambulance clinicians.**
REC reference:	**14/SC/0163**
Amendment number:	**PRoSAiC Amendment 1**
Amendment date:	**25 July 2014**
IRAS project ID:	**152449**

The above amendment was reviewed at the meeting of the Sub-Committee held on 06 August 2014.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

The Sub-Committee requested the following changes to the information sheet before approving the amendment.

1) In the third paragraph on the 1st page please replace the sentence 'These permissions have been granted' with 'These approvals have been given but with opportunity for patients to refuse participation.'

2) On the 2nd page, please remove the section 'Do I have to do anything?' as the same information is included in the section above this.

3) In the section 'What are the possible benefits of allowing us to use your medical record in this study?' please replace the words 'we will save more lives' with 'we should improve outcomes for patients.' – The Sub-Committee felt this was too bold a statement to make.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*

NHS
Health Research Authority

Copies of advertisement materials for research participants [Emergency Department Poster]	1.0	01 July 2014
Notice of Substantial Amendment (non-CTIMP)	PRoSAiC Amendment 1	25 July 2014
Participant information sheet (PIS)	1.1	06 August 2014
Research protocol or project proposal	1.1	27 June 2014

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

14/SC/0163: Please quote this number on all correspondence

Yours sincerely

pp

Professor Nigel Wellman
Chair

E-mail: nrescommittee.southcentral-oxfordc@nhs.net

Enclosures: *List of names and professions of members who took part in the review*

Copy to: *Mrs Gill Price, West Midlands Ambulance Service NHS Foundation Trust*
Mrs Jane Prewett Confidentiality Advise Team

A Research Ethics Committee established by the Health Research Authority

NHS
Health Research Authority

NRES Committee South Central - Oxford C

Attendance at Sub-Committee of the REC meeting on 06 August 2014

Committee Members:

Name	*Profession*	*Present*	*Notes*
Mr Barry Muir	Retired NHS Management Consultant	Yes	
Professor Nigel Wellman (Chair)	Professor of Health and Human Sciences	Yes	

Also in attendance:

Name	*Position (or reason for attending)*
Miss Lauren Allen	REC Manager

2.4 WMAS approval

West Midlands Ambulance Service NHS
NHS Trust

NIHR Clinical Research Network: West Midlands
Fourth Floor, West Wing (ACF40002)
University Hospitals Coventry & Warwickshire NHS Trust
University Hospital
Clifford Bridge Road
Coventry
CV2 2DX

9th April 2015

Mr Matthew Ward
Consultant Paramedic Emergency Care
West Midlands Ambulance Service NHS Foundation Trust
Millennium Point
Waterfront Business Park
Brierley Hill
West Midlands
DY5 1LX

Dear Mr Ward

Project Title: Prehospital Recognition of Sepsis by Ambulance Clinicians (PRoSAiC)
R&D Ref: 152449
REC Ref: 14/SC/0163

I am pleased to inform you that the R&D review of the above project is complete, and NHS permission has been granted for the study at West Midlands Ambulance Service NHS Foundation Trust. Your research activity is now covered by NHS indemnity as set out in HSG (96) 48, and your trial has been entered onto the Trust's database.

The permission has been granted on the basis described in the application form, protocol and supporting documentation. The documents reviewed were:

Document	Version	Date
R&D Form	152449/647765/14/366	-
SSI Form	152449/755614/6/648/241449/319938	-
REC Favourable Opinion Letter	-	06/08/2014
Confidentiality Advisory Group Approval	-	30/10/2014
Protocol	1.1	27/06/2014
Participant Information Sheet (PIS)	1.1	06/08/2014
Poster	1.0	01/07/2014
Data Flow Diagram	-	10/03/2014

All research must be managed in accordance with the requirements of the Department of Health's Research Governance Framework (RGF), to ICH-GCP standards (if applicable) and to NHS Trust policies and procedures. Permission is only granted for the activities agreed by the relevant authorities.

All amendments (including changes to the local research team and status of the project) need to be submitted to the REC and the R&D office in accordance with the guidance in IRAS. Any urgent safety measures required to protect research participants against immediate harm can be implemented immediately. You should notify the R&D Office within the same time frame as any other regulatory bodies.

It is your responsibility to keep the R&D Office and Sponsor informed of all Serious Adverse Events and to ensure that they are reported according to the Trust Clinical Incident policy, where required. All SAEs must be reported within the timeframes detailed within ICH-GCP statutory instruments and EU directives.

In order to ensure that research is carried out to the highest governance standards, the Trust employs the services of an external monitoring organisation to provide assurance. Your study may be randomly selected for audit at any time, and you must co-operate with the auditors. Action may be taken to suspend Trust approval if the research is not run in accordance with RGF or ICH-GCP standards, or following recommendations from the auditors.

You will be sent an annual progress report which must be completed in order to ensure that the information we hold on our database remains up to date, in line with RGF requirements.

I wish you well with your project. Please do not hesitate to contact me should you need any guidance or assistance.

Yours sincerely

Rachel Davis
Study Manager

Cc: Michael Smyth, Chief Investigator
Professor Gavin Perkins, Academic Supervisor
Professor Matthew Cooke, Academic Supervisor
Jane Prewett, Sponsor Representative (University of Warwick)
Gill Price, West Midlands Ambulance Service NHS Foundation Trust

2.5 UHNS approval

RESEARCH needs YOU®

University Hospitals of North Midlands NHS
NHS Trust

RESEARCH AND DEVELOPMENT DEPARTMENT
Academic Research Unit
Courtyard Annexe – C Block
Newcastle Road
Stoke-on-Trent
ST4 6QG
Telephone: 01782 675387
Fax: 01782 675399

Email: Darren.Clement@uhns.nhs.uk
research.governance@uhns.nhs.uk

Ref: DC/ds

1 April 2015

Dr Mark Ragoo
Consultant in Emergency Medicine
University Hospitals of North Midlands NHS Trust
Emergency Department
The Royal Stoke University Hospital
Newcastle Road
Stoke-on-Trent
ST4 6QG

Dear Dr Ragoo

Re: Study Title - PROSAIC
Chief Investigator: Mr Michael Smyth
Sponsor/Co-Sponsor: University of Warwick

I can confirm that the above project has been given NHS Permission for Research by the Research & Development Department for the University Hospitals of North Midlands NHS Trust and the details entered on to the R&D database.

I note that this research project has been approved by **Ethics Committee / REC ref: 14/SC/0163**

NHS permission for the above research has been granted on the basis described in the application form, protocol and supporting documentation. The documents reviewed were:

Document	Version Number	Date
Protocol	1.1	27 June 2014
Patient Information Sheet	1.1	06 August 2014
Emergency Department Poster	1.0	01 July 2014
Contract/Agreement		19 January 2015

The research sponsor or the Chief Investigator, or the local Principal Investigator at a research site, may take appropriate urgent safety measures in order to protect research participants against any immediate hazard to their health or safety. The R&D office should

R&D ID: UKCRN ID: CSP ID: REC REF:

be notified that such measures have been taken. The notification should also include the reasons why the measures were taken and the plan for further action. The R&D office should be notified within the same time frame of notifying the REC and any other regulatory bodies.

Approval by the R&D Dept therefore assumes that you have read, understand and agree to comply with the:

- Research Governance Framework (www.doh.gov.uk/research)
- Data Protection Act
- Health and Safety Act
- ICH Guidelines on good clinical practice
- All applicable Trust policies & procedures

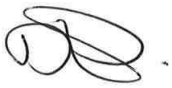
In line with these requirements may I draw your attention to the need for you to provide the following documentation/notifications to the R&D Department throughout the course of the study and that all amendments (including changes to the local research team) need to be submitted to R&D in accordance with guidance in IRAS:-

- Annual Progress Report Form (sent to you by this department)
- End of Study Declaration Form (available on IRAS website)
- Changes to study start and end dates
- Changes in study personnel

Please note that the NHS organisation is required to monitor research to ensure compliance with the Research Governance Framework and other legal and regulatory requirements. This will be achieved by random audit by our department.

I would like to take this opportunity to wish you well with your research. If you need any further advice or guidance please do not hesitate to contact us.

Yours sincerely



Dr Darren Clement
R&D Manager – University Hospitals of North Midlands NHS Trust

Cc **Chief Investigator – Mr Mike Smyth, Warwick Medical School**
Sponsor Representative – Mrs Jane Prewett, University of Warwick
SSI Authorisations – Dr John Oxtoby, Medical Director, UHNM
Jill Stacey – Acting Clinical Trials Lead Nurse, UHNM
Helen Grocott – Information Governance Manager, UHNM
Laura Longshaw – R&D Auditor & Monitor, UHNM

R&D ID: UKCRN ID: CSP ID: REC REF:

Appendix 3 Pragmatic variable adjustments

3.1 Adjustments to Age variable

Model A (GCS sum)

Age	Interval	ß$_i$	95 low	95 high	p-value
	below 40		reference		
	40-49	0.13	-0.56	0.82	0.71
	50-59	0.41	-0.19	1.01	0.18
	60-69	0.82	0.31	1.33	<0.001
	70-79	0.94	0.45	1.42	<0.001
	80-89	0.86	0.37	1.36	<0.001
	90 plus	1.34	0.78	1.89	<0.001

Model B (GCS verbal)

Age	Interval	ß$_i$	95 low	95 high	p-value
	below 40		reference		
	40-49	0.11	-0.58	0.81	0.75
	50-59	0.40	-0.20	1.00	0.19
	60-69	0.80	0.29	1.31	<0.001
	70-79	0.92	0.43	1.41	<0.001
	80-89	0.84	0.34	1.33	<0.001
	90 plus	1.31	0.76	1.87	<0.001

Model C (AVPU)

Age	Interval	ß$_i$	95 low	95 high	p-value
	below 40		reference		
	40-49	0.15	-0.54	0.84	0.68
	50-59	0.45	-0.15	1.05	0.14
	60-69	0.83	0.31	1.34	<0.001
	70-79	0.96	0.47	1.45	<0.001
	80-89	0.86	0.37	1.36	<0.001
	90 plus	1.32	0.77	1.88	<0.001

3.2 Adjustments to Resps variable

Model A (GCS sum)

Resps	Interval	ß$_i$	95 low	95 high	p-value
	below 10	-12.33	-579.16	554.49	0.97
	10-20		reference		
	21-25	0.53	0.20	0.87	<0.001
	26-30	0.89	0.57	1.22	<0.001
	31-35	1.01	0.57	1.44	<0.001
	36-40	1.27	0.90	1.65	<0.001
	41-50	1.66	1.10	2.21	<0.001
	51-60	1.78	1.00	2.56	<0.001
	60 plus	-11.57	-1964.57	1941.43	0.99

Model B (GCS verbal)

Resps	Interval	ß$_i$	95 low	95 high	p-value
	below 10	-12.47	-575.63	550.69	0.97
	10-20		reference		
	21-25	0.54	0.21	0.87	<0.001
	26-30	0.89	0.56	1.22	<0.001
	31-35	1.01	0.57	1.45	<0.001
	36-40	1.29	0.92	1.66	<0.001
	41-50	1.65	1.10	2.20	<0.001
	51-60	1.75	0.97	2.53	<0.001
	60 plus	-11.56	-1966.06	1942.94	0.99

Model C (AVPU)

Resps	Interval	ß$_i$	95 low	95 high	p-value
	below 10	-12.26	-579.25	554.74	0.97
	10-20		reference		
	21-25	0.55	0.22	0.88	<0.001
	26-30	0.91	0.59	1.24	<0.001
	31-35	1.02	0.58	1.45	<0.001
	36-40	1.31	0.94	1.68	<0.001
	41-50	1.66	1.12	2.20	<0.001
	51-60	1.77	0.99	2.55	<0.001
	60 plus	-11.51	-1966.16	1943.13	0.99

3.3 Adjustments to `Pulse` variable

Model A (GCS sum)

Pulse	Interval	ß$_i$	95 low	95 high	p-value
	below 60	-0.57	-1.42	0.28	0.19
	60-100		reference		
	101-110	0.62	0.29	0.94	<0.001
	111-120	0.64	0.30	0.98	<0.001
	121-130	0.96	0.59	1.34	<0.001
	131-140	0.89	0.44	1.33	<0.001
	141-160	1.59	1.16	2.03	<0.001
	161-170	0.46	-0.75	1.67	0.46
	171-180	0.10	-2.04	2.24	0.93
	180 plus	1.02	-0.20	2.25	0.10

Model B (GCS verbal)

Pulse	Interval	ß$_i$	95 low	95 high	p-value
	below 60	-0.58	-1.43	0.27	0.18
	60-100		reference		
	101-110	0.64	0.31	0.96	<0.001
	111-120	0.65	0.31	0.99	<0.001
	121-130	0.97	0.59	1.34	<0.001
	131-140	0.88	0.44	1.33	<0.001
	141-160	1.59	1.15	2.02	<0.001
	161-170	0.52	-0.67	1.71	0.39
	171-180	0.10	-2.04	2.24	0.93
	180 plus	1.01	-0.22	2.24	0.11

Model C (AVPU)

Pulse	Interval	ß$_i$	95 low	95 high	p-value
	below 60	-0.57	-1.42	0.27	0.19
	60-100		reference		
	101-110	0.62	0.30	0.95	<0.001
	111-120	0.65	0.32	0.99	<0.001
	121-130	0.98	0.61	1.35	<0.001
	131-140	0.91	0.47	1.35	<0.001
	141-160	1.64	1.21	2.07	<0.001
	161-170	0.79	-0.32	1.90	0.16
	171-180	0.15	-1.99	2.28	0.89
	180 plus	1.07	-0.14	2.29	0.08

References

1. Horn K. Evolving strategies in the treatment of sepsis and systemic inflammatory response syndrome (SIRS). *Quarterly Journal of Medicine* 1998;91(4):265--77.
2. Cohen J. The immunopathogenesis of sepsis. *Nature* 2002;420(6917):885--91.
3. Martin GS, Mannino DM, Eaton S, et al. The epidemiology of sepsis in the United States from 1979 through 2000. *New England Journal of Medicine* 2003;348(16):1546--54.
4. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *Journal of the American Medical Association* 2016;315(8):801--10.
5. UK Sepsis Trust. Toolkit: Emergency Department management of Sepsis in adults and young people over 12 years- 2016. 2016
. http://sepsistrust.org/wp-content/uploads/2016/07/ED-toolkit-2016-Final-1.pdf.
6. Stearns-Kurosawa DJ, Osuchowski MF, Valentine C, et al. The pathogenesis of sepsis. *Annual Review of Pathology: Mechanisms of Disease* 2011;6:19--48.
7. Cinel I, Opal SM. Molecular biology of inflammation and sepsis: a primer. *Critical Care Medicine* 2009;37(1):291--304.
8. Villar J, Maca-Meyer N, Pérez-Méndez L, et al. Bench-to-bedside review: Understanding genetic predisposition to sepsis. *Critical Care* 2004;8(3):180. doi: 10.1186/cc2863
9. Rooney CI, Smith SK. Implementation of ICD-10 for mortality data in England and Wales from January 2001. *Health Statistics Quarterly* 2000(8):41-51.
10. McPherson D, Griffiths C, Williams M, et al. Sepsis-associated mortality in England: an analysis of multiple cause of death data from 2001 to 2010. *BMJ Open* 2013;3(8) doi: 10.1136/bmjopen-2013-002586
11. Burchardi H, Schneider H. Economic aspects of severe sepsis: a review of intensive care unit costs, cost of illness and cost effectiveness of therapy. *PharmacoEconomics* 2004;22(12):793-813.
12. Hex N, Retzler J, Bartlett C, et al. The cost of sepsis care in the UK. 2017. http://sepsistrust.org/wp-content/uploads/2017/02/YHEC-Sepsis-Report-17.02.17-FINAL.pdf.
13. Bone RC, Balk RA, Cerra FB, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992;101(6):1644-55.
14. Bone RC, Sibbald WJ, Sprung CL. The ACCP-SCCM consensus conference on sepsis and organ failure. *Chest* 1992;101(6):1481-3.
15. Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Critical Care Medicine* 2013;41(2):580-637. doi: 10.1097/CCM.0b013e31827e83af
16. Churpek MM, Zadravecz FJ, Winslow C, et al. Incidence and Prognostic Value of the Systemic Inflammatory Response Syndrome and Organ Dysfunctions in Ward Patients. *American Journal of Respiratory and Critical Care Medicine* 2015 doi: 10.1164/rccm.201502-0275OC
17. Kaukonen KM, Bailey M, Pilcher D, et al. Systemic inflammatory response syndrome criteria in defining severe sepsis. *New England Journal of Medicine* 2015;372(17):1629-38. doi: 10.1056/NEJMoa1415236

18. Vincent JL. Dear SIRS, I'm sorry to say that I don't like you. *Critical Care Medicine* 1997;25(2):372-4.
19. Vincent JL, Beumier M. Diagnostic and prognostic markers in sepsis. *Expert Review Antibiotics Infection Therapies* 2013;11(3):265-75. doi: 10.1586/eri.13.9
20. Vincent JL, Opal SM, Marshall JC, et al. Sepsis definitions: time for change. *Lancet* 2013;381(9868):774-5. doi: 10.1016/S0140-6736(12)61815-7
21. National Institute for Health Care Excellence. Sepsis: recognition, diagnosis and early management. [ng51]: NICE, 2016.
22. Toolkit: Emergency Department management of Sepsis in adults and young people over 12 years-2016 [program]: UK Sepsis Trust, 2016.
23. Guerra WF, Mayfield TR, Meyers MS, et al. Early detection and treatment of patients with severe sepsis by prehospital personnel. *Journal of Emergency Medicine* 2013;44(6):1116-25. doi: http://dx.doi.org/10.1016/j.jemermed.2012.11.003
24. Suffoletto B, Frisch A, Prabhu A, et al. Prediction of serious infection during prehospital emergency care. *Prehospital Emergency Care* 2011;15(3):325-30. doi: http://dx.doi.org/10.3109/10903127.2011.561411
25. Seymour CW, Rea TD, Kahn JM, et al. Severe sepsis in pre-hospital emergency care: analysis of incidence, care, and outcome. *American Journal of Respiratory & Critical Care Medicine* 2012;186(12):1264-71. doi: http://dx.doi.org/10.1164/rccm.201204-0713OC
26. Ibrahim I, Jacobs IG. Can the characteristics of emergency department attendances predict poor hospital outcomes in patients with sepsis? *Singapore Medical Journal* 2013;54(11):634-8.
27. Studnek JR, Artho MR, Garner CL, Jr., et al. The impact of emergency medical services on the ED care of severe sepsis. *American Journal of Emergency Medicine* 2012;30(1):51-6. doi: http://dx.doi.org/10.1016/j.ajem.2010.09.015
28. Wang HE, Weaver MD, Shapiro NI, et al. Opportunities for Emergency Medical Services care of sepsis. *Resuscitation* 2010;81(2):193-7. doi: http://dx.doi.org/10.1016/j.resuscitation.2009.11.008
29. Femling J, Weiss S, Hauswald E, et al. EMS patients and walk-in patients presenting with severe sepsis: differences in management and outcome. *Southern Medical Journal* 2014;107(12):751-6. doi: 10.14423/SMJ.0000000000000206
30. Gray A, Ward K, Lees F, et al. The epidemiology of adults with severe sepsis and septic shock in Scottish emergency departments. *Emergency Medicine Journal* 2013;30(5):397-401. doi: http://dx.doi.org/10.1136/emermed-2012-201361
31. Erwin A, Salzman J, Wewerka S, et al. Can paramedics accurately diagnose sepsis and severe sepsis in the field? *Academic Emergency Medicine* 2011;1:S23-S24. doi: http://dx.doi.org/10.1111/j.1553-2712.2011.01073.x
32. Guerra WF, Mayfield TR, Meyers MS, et al. Early detection and treatment of patients with severe sepsis by prehospital personnel. *Journal of Emergency Medicine* 2013;44(6):1116-25. doi: 10.1016/j.jemermed.2012.11.003
33. McClelland G, Jones J. A pilot study exploring the accuracy of pre-hospital sepsis recognition in the North East Ambulance Service. *Journal of Paramedic Practice* 2015;7(9):459-65.
34. Shiuh T, Sweeney T, Rupp R, et al. An emergency medical services sepsis protocol with point-of-care lactate accurately identifies out-of-hospital patients with severe infection and sepsis. *Annals of Emergency Medicine*

2012;1):S44. doi: http://dx.doi.org/10.1016/j.annemergmed.2012.06.097
35. Travers A, Green R, Cain E, et al. Can paramedics diagnose sepsis in the prehospital setting? A feasibility study. *Canadian Journal of Emergency Medicine* 2013;15:S55.
36. Wallgren UM, Castren M, Svensson AE, et al. Identification of adult septic patients in the prehospital setting: a comparison of two screening tools and clinical judgment. *European Journal of Emergency Medicine* 2014;21(4):260-5. doi: 10.1097/MEJ.0000000000000084
37. Baez AA, Hanudel P, Perez MT, et al. Prehospital Sepsis Project (PSP): knowledge and attitudes of United States advanced out-of-hospital care providers. *Prehospital & Disaster Medicine* 2013;28(2):104-6. doi: http://dx.doi.org/10.1017/S1049023X12001744
38. Shime N. A Survey of the Competency of Ambulance Service Personnel in the Diagnosis and Management of Sepsis. *Journal of Emergency Medicine* 2015 doi: 10.1016/j.jemermed.2014.12.066
39. Smyth MA, Brace-McDonnell SJ, Perkins GD. Identification of adults with sepsis in the prehospital environment: a systematic review. *BMJ Open* 2016;6(8):e011218. doi: 10.1136/bmjopen-2016-011218
40. Robson WN, T.; Daniels R. Sepsis: a need for prehospital intervention? *Emergency Medicine Journal* 2009;26(7):535-8. doi: http://dx.doi.org/10.1136/emj.2008.064469
41. NHS England. Patient Safety alert - resources to support the prompt recognition of sepsis and the rapid initiation of treatment 2014 [Available from: https://www.england.nhs.uk/2014/09/02/psa-sepsis/ accessed 10 September 2014 2014.
42. Daniels R, Nutbeam T, McNamara G, et al. The sepsis six and the severe sepsis resuscitation bundle: a prospective observational cohort study. *Emergency Medicine Journal* 2011;28(6):507-12. doi: http://dx.doi.org/10.1136/emj.2010.095067
43. Physicians RCo. Acute care toolkit 9: Sepsis. 2015. https://www.rcplondon.ac.uk/file/1182/download?token=fQdN_6kB (accessed September).
44. Association of Ambulance Services Chief Executives. UK Ambulance Services Clinical Practice Guidelines: Class Professional Publishing 2016.
45. Hanudel P, Wilcox S, Cadin E, et al. Prevalence of cryptic shock in a cohort of out-of-hospital sepsis patients: an argument for out of hospital point-of-care lactate. *Annals of emergency medicine* 2008;51(1):s487-89.
46. Hokanson J, Boland L, Olson T, et al. Use of lactate meters and temporal artery thermometers by paramedics to aid in the out-of-hospital recognition of sepsis: A pilot study. *Annals of Emergency Medicine* 2012;1):S43. doi: http://dx.doi.org/10.1016/j.annemergmed.2012.06.095
47. Jansen TC, van Bommel J, Mulder PG, et al. The prognostic value of blood lactate levels relative to that of vital signs in the pre-hospital setting: a pilot study. *Critical Care* 2008;12(6) doi: 10.1186/cc7159
48. Tobias AZ, Guyette FX, Seymour CW, et al. Pre-resuscitation lactate and hospital mortality in prehospital patients. *Prehospital Emergency Care* 2014;18(3):321-7. doi: 10.3109/10903127.2013.869645
49. van Beest PA, Mulder PJ, Oetomo SB, et al. Measurement of lactate in a prehospital setting is related to outcome. *European Journal of Emergency Medicine* 2009;16(6):318-22. doi: 10.1097/MEJ.0b013e32832dbe54
50. Bion JF, Wilson IH, Taylor PA. Transporting critically ill patients by ambulance: audit by sickness scoring. *British Medical Journal* 1988;296(6616):170.

51. PONS PT, MOORE EE, CUSICK JM, et al. Prehospital Venous Access in an Urban Paramedic System-A Prospective On-scene Analysis. *Journal of Trauma and Acute Care Surgery* 1988;28(10):1460-63.
52. Prause G, Ratzenhofer-Komenda B, Offner A, et al. Prehospital point of care testing of blood gases and electrolytes — an evaluation of IRMA. *Critical Care* 1997;1(2):79-83.
53. Band RA, Gaieski DF, Hylton JH, et al. Arriving by emergency medical services improves time to treatment endpoints for patients with severe sepsis or septic shock. *Academic Emergency Medicine* 2011;18(9):934-40. doi: http://dx.doi.org/10.1111/j.1553-2712.2011.01145.x
54. Bayer O, Schwarzkopf D, Stumme C, et al. An Early Warning Scoring System to Identify Septic Patients in the Prehospital Setting: The PRESEP Score. *Academic Emergency Medicine* 2015;22(7):868-71. doi: 10.1111/acem.12707
55. Chamberlain D. Prehospital administered intravenous antimicrobial protocol for septic shock: A prospective randomized clinical trial. *Critical Care* 2009;13:S130-S31. doi: http://dx.doi.org/10.1186/cc7481
56. Seymour CW, Cooke CR, Heckbert SR, et al. Prehospital intravenous access and fluid resuscitation in severe sepsis: an observational cohort study. *Critical Care* 2014;18(5):533. doi: 10.1186/s13054-014-0533-x
57. Moons KG, Royston P, Vergouwe Y, et al. Prognosis and prognostic research: what, why, and how? *British Medical Journal* 2009;338:b375.
58. Labarère J, Bertrand R, Fine MJ. How to derive and validate clinical prediction models for use in intensive care medicine. *Intensive Care Medicine* 2014;40(4):513--27.
59. Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating: Springer Science \& Business Media 2008.
60. Toll D, Janssen K, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. *Journal of Clinical Epidemiology* 2008;61(11):1085-94.
61. Heckerling PS, Tape TG, Wigton RS, et al. Clinical prediction rule for pulmonary infiltrates. *Annals of Internal Medicine* 1990;113(9):664-70.
62. Wasson JH, Sox HC, Neff RK, et al. Clinical prediction rules: applications and methodological standards. *New England Journal of Medicine* 1985;313(13):793--99.
63. Feinstein AR. Clinimetric perspectives. *Journal of Chronic Diseases* 1987;40(6):635-40.
64. Laupacis A, Sekar N, others. Clinical prediction rules: a review and suggested modifications of methodological standards. *Journal of the American Medical Association* 1997;277(6):488--94.
65. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Annals of Emergency Medicine* 1999;33(4):437--47.
66. Moons KG, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *British Medical Journal* 2009;338:b606.
67. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012:heartjnl--2011.
68. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker. *Heart* 2012;98(9):683--90.
69. Royston P, Moons KG, Altman DG, et al. Prognosis and prognostic research: developing a prognostic model. *British Medical Journal* 2009;338:b604.

70. Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *Public Library of Science Medicine* 2012;9(5):e1001221.
71. Harrell F. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer 2015.
72. Padkin A, Goldfrad C, Brady AR, et al. Epidemiology of severe sepsis occurring in the first 24 hrs in intensive care units in England, Wales, and Northern Ireland. *Critical Care Medicine* 2003;31(9):2332-8. doi: 10.1097/01.CCM.0000085141.75513.2B
73. Gaieski DF, Mikkelsen ME, Band RA, et al. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Critical Care Medicine* 2010;38(4):1045-53. doi: http://dx.doi.org/10.1097/CCM.0b013e3181cc4824
74. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine* 2006;34(6):1589-96. doi: 10.1097/01.CCM.0000217961.75225.E9
75. Mouncey PR, Osborn TM, Power GS, et al. Trial of early, goal-directed resuscitation for septic shock. *New England Journal of Medicine* 2015;372(14):1301-11. doi: 10.1056/NEJMoa1500896
76. Peake SL, Delaney A, Bailey M, et al. Goal-directed resuscitation for patients with early septic shock. *New England Journal of Medicine* 2014;371(16):1496-506. doi: 10.1056/NEJMoa1404380
77. Stolmeijer R, ter Maaten JC, Zijlstra JG, et al. Oxygen therapy for sepsis patients in the emergency department: a little less? *European Journal of Emergency Medicine* 2014;21(3):233-5. doi: http://dx.doi.org/10.1097/MEJ.0b013e328361c6c7
78. Yealy DM, Kellum JA, Huang DT, et al. A randomized trial of protocol-based care for early septic shock. *New England Journal of Medicine* 2014;370(18):1683-93. doi: 10.1056/NEJMoa1401602
79. Cannon CM, Holthaus CV, Zubrow MT, et al. The GENESIS project (GENeralized Early Sepsis Intervention Strategies): a multicenter quality improvement collaborative. *J Intensive Care Med* 2013;28(6):355-68. doi: http://dx.doi.org/10.1177/0885066612453025
80. Keegan J, Wira CR, 3rd. Early identification and management of patients with severe sepsis and septic shock in the emergency department. *Emergency Medicine Clinics of North America* 2014;32(4):759-76. doi: http://dx.doi.org/10.1016/j.emc.2014.07.002
81. Wira CR, Dodge K, Sather J, et al. Meta-analysis of protocolized goal-directed hemodynamic optimization for the management of severe sepsis and septic shock in the Emergency Department. *Western Journal of Emergency Medicine* 2014;15(1):51-9. doi: 10.5811/westjem.2013.7.6828
82. Abdullah AR, Smith EE, Biddinger PD, et al. Advance hospital notification by EMS in acute stroke is associated with shorter door-to-computed tomography time and increased likelihood of administration of tissue-plasminogen activator. *Prehospital Emergency Care* 2008;12(4):426-31. doi: 10.1080/10903120802290828
83. Acker JE, 3rd, Pancioli AM, Crocco TJ, et al. Implementation strategies for emergency medical services within stroke systems of care: a policy statement from the American Heart Association/American Stroke Association Expert Panel on Emergency Medical Services Systems and the Stroke Council. *Stroke; a journal of cerebral circulation* 2007;38(11):3097-115. doi: 10.1161/STROKEAHA.107.186094

84. American College of Emergency Physicians. Trauma care systems development, evaluation, and funding. *Annals of Emergency Medicine* 2006;47(3):306. doi: 10.1016/j.annemergmed.2005.11.015
85. American College of Emergency Physicians. Trauma care systems development, evaluation, and funding. Policy statement. *Ann Emerg Med* 2012;60(2):249-50. doi: 10.1016/j.annemergmed.2012.04.027
86. Blackwell T, Kellam JF, Thomason M. Trauma care systems in the United States. *Injury* 2003;34(9):735-9.
87. Cooper G, Laskowski-Jones L. Development of trauma care systems. *Prehospital Emergency Care* 2006;10(3):328-31. doi: 10.1080/10903120600728755
88. Jollis JG, Roettig ML, Aluko AO, et al. Implementation of a statewide system for coronary reperfusion for ST-segment elevation myocardial infarction. *Journal of the American Medical Association* 2007;298(20):2371-80. doi: 10.1001/jama.298.20.joc70124
89. Moyer P, Ornato JP, Brady WJ, Jr., et al. Development of systems of care for ST-elevation myocardial infarction patients: the emergency medical services and emergency department perspective. *Circulation* 2007;116(2):e43-8. doi: 10.1161/CIRCULATIONAHA.107.184047
90. Nathens AB, Jurkovich GJ, Cummings P, et al. The effect of organized systems of trauma care on motor vehicle crash mortality. *Journal of the American Medical Association* 2000;283(15):1990-4.
91. Strauss DG, Sprague PQ, Underhill K, et al. Paramedic transtelephonic communication to cardiologist of clinical and electrocardiographic assessment for rapid reperfusion of ST-elevation myocardial infarction. *Journal of Electrocardiology* 2007;40(3):265-70. doi: 10.1016/j.jelectrocard.2006.11.006
92. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* 2009;62(10):1006-12. doi: 10.1016/j.jclinepi.2009.06.005
93. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal* 2008;336:924-26. doi: 10.1136/bmj.39489.470347.AD
94. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *Journal of Clinical Epidemiology* 2011;64(4):407-15. doi: 10.1016/j.jclinepi.2010.07.017
95. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *Journal of Clinical Epidemiology* 2011;64(12):1303-10. doi: 10.1016/j.jclinepi.2011.04.014
96. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. *Journal of Clinical Epidemiology* 2011;64(12):1294-302. doi: 10.1016/j.jclinepi.2011.03.017
97. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 2011;64(4):383-94. doi: 10.1016/j.jclinepi.2010.04.026
98. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. *Journal of Clinical Epidemiology* 2011;64(12):1277-82. doi: 10.1016/j.jclinepi.2011.01.011
99. Spiegel R, Opic P, Semmlack S, et al. Tackling submission and publication bias. *British Medical Journal* 2017;358
100. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *Journal of Clinical Epidemiology* 2011;64(12):1311-6. doi: 10.1016/j.jclinepi.2011.06.004
101. Baez AA, Hanudel P, Wilcox SR. The Prehospital Sepsis Project: Out-of-Hospital Physiologic Predictors of Sepsis Outcomes. *Prehospital &*

Disaster Medicine 2013;28(6):632-35. doi: 10.1017/S1049023X1300890X

102. Bayer O, Stumme C, Schneider K, et al. Evaluation of 1 year prehospital emergency sepsis care: An observational study. *Infection, Supplement* 2013;1):S66. doi: http://dx.doi.org/10.1007/s15010-013-0513-0
103. Erwin A, Salzman J, Wewerka S, et al. Can paramedics accurately diagnose sepsis and severe sepsis in the field? *Academic Emergency Medicine* 2011;1):S23-S24. doi: http://dx.doi.org/10.1111/j.1553-2712.2011.01073.x
104. Femling JK, Weiss SJ, Fullerton L, et al. EMS sepsis patients are different from walk-in sepsis patients. *Annals of Emergency Medicine* 2012;1):S120. doi: http://dx.doi.org/10.1016/j.annemergmed.2012.06.316
105. Milzman D, Moynihan M, Phillips C. Early identification of sepsis patients using shock index and temperature in prehospital setting predicts increased hospitalization and survival. *Crit Care Med* 2012;1):139. doi: http://dx.doi.org/10.1097/01.ccm.0000425605.04623.4b
106. Polito C, Isakov A, Yancey A, et al. Detection of severe sepsis in the pre-hospital setting: How are we doing? *Critical Care Medicine* 2014;1):A1603. doi: http://dx.doi.org/10.1097/01.ccm.0000458506.83088.d1
107. Seymour CW, Kahn JM, Cooke CR, et al. Prediction of critical illness during out-of-hospital emergency care. *Journal of the American Medical Association* 2010;304(7):747-54. doi: http://dx.doi.org/10.1001/jama.2010.1140
108. Polito CC, Isakov A, Yancey n, Arthur H, et al. Prehospital recognition of severe sepsis: development and validation of a novel EMS screening tool. *American Journal of Emergency Medicine* 2015;33(9):1119-25. doi: 10.1016/j.ajem.2015.04.024
109. Milzman DM, M.; Phillips, C. Early identification of sepsis patients using shock index and temperature in prehospital setting predicts increased hospitalization and survival. *Critical Care Medicine* 2012;1):139. doi: http://dx.doi.org/10.1097/01.ccm.0000425605.04623.4b
110. Stenhouse C, Coates S, Tivey M, et al. Prospective evaluation of a modified Early Warning Score to aid earlier detection of patients developing critical illness on a general surgical ward. *British Journal of Anaesthesia* 2000;84(5):663.
111. Seymour CW, Cooke CR, Mikkelsen ME, et al. Out-of-hospital fluid in severe sepsis: effect on early resuscitation in the emergency department. *Prehospital Emergency Care* 2010;14(2):145-52. doi: http://dx.doi.org/10.3109/10903120903524997
112. Báez AA, Hanudel P, Perez MT, et al. Prehospital Sepsis Project (PSP): knowledge and attitudes of United States advanced out-of-hospital care providers. *Prehospital and disaster medicine* 2013;28(02):104--06.
113. Tusgul S, Carron PN, Yersin B, et al. Low sensitivity of qSOFA, SIRS criteria and sepsis definition to identify infected patients at risk of complication in the prehospital setting and at the emergency department triage. *Scand J Trauma Resusc Emerg Med* 2017;25(1):108. doi: 10.1186/s13049-017-0449-y [published Online First: 2017/11/05]
114. Walchok JG, Pirrallo RG, Furmanek D, et al. Paramedic-Initiated CMS Sepsis Core Measure Bundle Prior to Hospital Arrival: A Stepwise Approach. *Prehosp Emerg Care* 2017;21(3):291-300. doi: 10.1080/10903127.2016.1254694 [published Online First: 2016/12/06]

115. Hunter CL, Silvestri S, Ralls G, et al. A prehospital screening tool utilizing end-tidal carbon dioxide predicts sepsis and severe sepsis. *Am J Emerg Med* 2016;34(5):813-9. doi: 10.1016/j.ajem.2016.01.017 [published Online First: 2016/02/18]
116. Boland LL, Hokanson JS, Fernstrom KM, et al. Prehospital Lactate Measurement by Emergency Medical Services in Patients Meeting Sepsis Criteria. *West J Emerg Med* 2016;17(5):648-55. doi: 10.5811/westjem.2016.6.30233 [published Online First: 2016/09/15]
117. Dorsett M, Kroll M, Smith CS, et al. qSOFA Has Poor Sensitivity for Prehospital Identification of Severe Sepsis and Septic Shock. *Prehosp Emerg Care* 2017;21(4):489-97. doi: 10.1080/10903127.2016.1274348 [published Online First: 2017/01/26]
118. Jouffroy R, Saade A, Ellouze S, et al. Prehospital triage of septic patients at the SAMU regulation: Comparison of qSOFA, MRST, MEWS and PRESEP scores. *Am J Emerg Med* 2017 doi: 10.1016/j.ajem.2017.10.030 [published Online First: 2017/10/24]
119. Axelsson C, Herlitz J, Karlsson A, et al. The Early Chain of Care in Patients with Bacteraemia with the Emphasis on the Prehospital Setting. *Prehosp Disaster Med* 2016;31(3):272-7. doi: 10.1017/S1049023X16000339 [published Online First: 2016/03/31]
120. Carberry M, Harden J. A collaborative improvement project by an NHS Emergency Department and Scottish Ambulance Paramedics to improve the identification and delivery of sepsis 6. *BMJ Qual Improv Rep* 2016;5(1) doi: 10.1136/bmjquality.u212670.w5049 [published Online First: 2016/10/19]
121. Alam N, Oskam E, Stassen PM, et al. Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. *Lancet Respir Med* 2018;6(1):40-50. doi: 10.1016/S2213-2600(17)30469-1 [published Online First: 2017/12/03]
122. Lane D, Ichelson RI, Drennan IR, et al. Prehospital management and identification of sepsis by emergency medical services: a systematic review. *Emerg Med J* 2016;33(6):408-13. doi: 10.1136/emermed-2015-205261 [published Online First: 2016/02/13]
123. West Midlands Ambulance Service. Operational Notice: Documentation of Patient Care: West Midlands Ambulance Service, 2012.
124. Müller H, Freytag J-C. Problems, methods, and challenges in comprehensive data cleansing: Professoren des Inst. Für Informatik 2005.
125. Rahm E, Do HH. Data cleaning: Problems and current approaches. *IEEE Data Enginerring Bulletin* 2000;23(4):3--13.
126. Chu X, Ilyas IF, Krishnan S, et al. Data cleaning: Overview and emerging challenges. Proceedings of the 2016 International Conference on Management of Data, 2016:2201--06.
127. van Buuren S. Flexible imputation of missing data: CRC press 2012.
128. Knaus WA, Zimmerman JE, Wagner DP, et al. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Critical Care Medicine* 1981;9(8):591-97.
129. Christen P. Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection: Springer Science \& Business Media 2012.
130. Harron K, Goldstein H, Dibben C. Methodological developments in data linkage: John Wiley \& Sons 2015.
131. Dusetzina SB, Tyree S, Meyer A, et al. Linking data for health services research: a framework and instructional guide: Agency for Healthcare Research and Quality (US), Rockville (MD) 2014.

132. Sariyar M, Borg A. The RecordLinkage package: Detecting errors in data. *The R Journal* 2010;2(2):61--67.
133. Grannis SJ, Overhage JM, McDonald CJ. Analysis of identifier performance using a deterministic linkage algorithm. *Proceedings of the AMIA Symposium* 2002:305--09.
134. Fellegi IP, Sunter AB. A theory for record linkage. *Journal of the American Statistical Association* 1969;64(328):1183--210.
135. Zhu VJ, Overhage MJ, Egg J, et al. An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. *Journal of the American Medical Informatics Association* 2009;16(5):738--45.
136. Porter EH, Winkler WE, others. Approximate string comparison and its effect on an advanced record linkage system. Advanced record linkage system US Bureau of the Census, Research Report, 1997.
137. Winkler WE, Thibaudeau Y. An application of the Fellegi-Sunter model of record linkage to the 1990 US decennial census. *US Bureau of the Census* 1991:1--22.
138. Elmagarmid AK, Ipeirotis PG, Verykios VS. Duplicate record detection: A survey. *IEEE Transactions on knowledge and data engineering* 2007;19(1)
139. Newgard CD. Validation of probabilistic linkage to match de-identified ambulance records to a state trauma registry. *Academic Emergency Medicine* 2006;13(1):69--75.
140. Downing A, Wilson R, Cooke M. Linkage of ambulance service and accident and emergency department data: a study of assault patients in the west midlands region of the UK. *Injury* 2005;36(6):738--44.
141. Mears GD, Rosamond WD, Lohmeier C, et al. A link to improve stroke patient care: a successful linkage between a statewide emergency medical services data system and a stroke registry. *Academic Emergency Medicine* 2010;17(12):1398--404.
142. Haukoos JS, Newgard CD. Advanced statistics: missing data in clinical research---part 1: an introduction and conceptual framework. *Academic Emergency Medicine* 2007;14(7):662--68.
143. Rubin DB. Inference and missing data. *Biometrika* 1976:581--92.
144. Allison PD. Missing data: Quantitative applications in the social sciences. *British Journal of Mathematical and Statistical Psychology* 2002;55(1):193--96.
145. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological Methods* 2002;7(2):147.
146. Acock AC. Working with missing values. *Journal of Marriage and Family* 2005;67(4):1012--28.
147. Graham JW. Missing data analysis: Making it work in the real world. *Annual Review of Psychology* 2009;60:549--76.
148. Soley-Bori M. Dealing with missing data: Key assumptions and methods for applied analysis. *Boston University* 2013
149. Wothke W. Longitudinal and multigroup modeling with missing data.: Lawrence Erlbaum Associates Publishers 2000.
150. Enders CK. Applied missing data analysis: Guilford Press 2010.
151. Nakai M, Ke W. Review of the methods for handling missing data in longitudinal data analysis. *International Journal of Mathematical Analysis* 2011;5(1):1--13.
152. Eekhout I, de Boer RM, Twisk JW, et al. Missing data: a systematic review of how they are reported and handled. *Epidemiology* 2012;23(5):729--32.
153. Baraldi AN, Enders CK. An introduction to modern missing data analyses. *Journal of School Psychology* 2010;48(1):5--37.

154. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials* 2004;1(4):368--76.
155. Durrant GB, others. Imputation methods for handling item-nonresponse in the social sciences: a methodological review. *ESRC National Centre for Research Methods and Southampton Statistical Sciences Research Institute NCRM Methods Review Papers NCRM/002* 2005
156. Roth PL. Missing data: A conceptual review for applied psychologists. *Personnel Psychology* 1994;47(3):537--60.
157. Fox-Wasylyshyn SM, El-Masri MM. Handling missing data in self-report measures. *Research in Nursing & Health* 2005;28(6):488--95.
158. Brown ML, Kros JF. Data mining and the impact of missing data. *Industrial Management & Data Systems* 2003;103(8):611--21.
159. Fayers PM, Machin D. Missing data. *Quality of Life: Assessment, Analysis and Interpretation* 2000:224--47.
160. Eekhout I, de Vet HC, Twisk JW, et al. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *Journal of Clinical Epidemiology* 2014;67(3):335--42.
161. Schafer JL. Multiple imputation: a primer. *Statistical Methods in Medical Research* 1999;8(1):3--15.
162. Pigott TD. A review of methods for missing data. *Educational Research and Evaluation* 2001;7(4):353--83.
163. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal* 2009;338:b2393.
164. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons 2004.
165. van Buuren S, Groothuis-Oudshoorn K, Robitzsch A, et al. Package `mice'. *[Computer software] Retrieved from:* http://cran/ *r-project org/web/packages/mice/mice pdf* 2015
166. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science* 2007;8(3):206--13.
167. Bodner TE. What improves with increased missing data imputations? *Structural Equation Modeling* 2008;15(4):651--75.
168. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine* 2011;30(4):377--99.
169. Gelman A, Hill J, Su Y-S, et al. Package `mi'. 2015
170. Groenewoudt M, Roest AA, Leijten FMM, et al. Septic patients arriving with emergency medical services: A seriously ill population. *European Journal of Emergency Medicine* 2014;21(5):330-35. doi: http://dx.doi.org/10.1097/MEJ.0000000000000091
171. van der Wekken LC, Alam N, Holleman F, et al. Epidemiology of Sepsis and Its Recognition by Emergency Medical Services Personnel in the Netherlands. *Prehospital Emergency Care* 2015 doi: 10.3109/10903127.2015.1037476
172. Roest AA, Stoffers J, Pijpers E, et al. Ambulance patients with nondocumented sepsis have a high mortality risk: a retrospective study. *European Journal of Emergency Medicine* 2015 doi: 10.1097/mej.0000000000000302
173. Wang HE, Shapiro NI, Angus DC, et al. National estimates of severe sepsis in United States emergency departments. *Critical Care Medicine* 2007;35(8):1928-36. doi: 10.1097/01.CCM.0000277043.85378.C1

174. Smyth MA, Brace-McDonnell SJ, Perkins GD. Impact of Prehospital Care on Outcomes in Sepsis: A Systematic Review. *Western Journal of Emergency Medicine* 2016;17(4):427-37. doi: 10.5811/westjem.2016.5.30172
175. Harrell F. Regression coefficients and scoring rules. *Journal of Clinical Epidemiology* 1996;49(7):819.
176. Harrell FE, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 1996;15:361-87.
177. Harrell FE, Margolis PA, Gove S, et al. Development of a clinical prediction model for an ordinal outcome: the World Health Organization Multicentre Study of Clinical Signs and Etiological agents of Pneumonia, Sepsis and Meningitis in Young Infants. *Statistics in Medicine* 1998;17(8):909--44.
178. Vergouwe Y, Steyerberg EW, Eijkemans MJ, et al. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *Journal of Clinical Epidemiology* 2005;58(5):475-83.
179. Tu YK, Kellett M, Clerehugh V, et al. Problems of correlations between explanatory variables in multiple regression analyses in the dental literature. *British Dental Journal* 2005;199(7):457-61.
180. Altman DG. Statistics in medical journals. *Statistics in Medicine* 1982;1:59-71.
181. Altman DG. Statistics in medical journals: developments in the 1980s. *Statistics in Medicine* 1991;10:1897-913.
182. Tu Y, Gilthorpe MS, Griffiths GS. Is reduction of pocket probing depth correlated with the baseline value or is it 'mathematical coupling'? *Journal of Dental Research* 2002;81:722-26.
183. Tu YK, Clerehugh V, Gilthorpe MS. Ratio variables in regression analysis can give rise to spurious results: a lesson from guided tissue regeneration. *J Dent* 2004;32:143-51.
184. Tu YK, Maddick IH, Griffiths GS, et al. Mathematical coupling still undermines the statistical assessment of clinical research: illustration from the treatment of guided tissue regeneration. *Journal of Dentistry* 2004;32:133-42.
185. Everitt BS. The Cambridge dictionary of statistics: Cambridge University Press 2006.
186. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine* 2006;25(1):127-41.
187. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *New England Journal of Medicine* 1997;336(4):243-50.
188. Aujesky D, Obrosky DS, Stone RA, et al. Derivation and validation of a prognostic model for pulmonary embolism. *American Journal of Respiratory and Critical Care Medicine* 2005;172(8):1041-46.
189. Lemeshow S, Hosmer Jr DW. A review of goodness of fit statistics for use in the development of logistic regression models. *American Journal of Epidemiology* 1982;115(1):92-106.
190. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons 2013.
191. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Critical Care Medicine* 2007;35(9):2052-6. doi: 10.1097/01.CCM.0000275267.64078.B0

192. Miller ME, Langefeld CD, Tierney WM, et al. Validation of probabilistic predictions. *Medical Decision Making* 1993;13(1):49-58. doi: 10.1177/0272989X9301300107
193. Austin PC, Steyerberg EW. Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. *BMC Medical Research Methodology* 2012;12:82-82. doi: 10.1186/1471-2288-12-82
194. Steyerberg EW, Harrell FE, Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology* 2001;54(8):774-81.
195. Grzybowski M, Younger JG. Statistical methodology: III. Receiver operating characteristic (ROC) curves. *Academic Emergency Medicine* 1997;4(8):818-26.
196. Allgower M, Burri C. Shock index. *Deutsche Medizinische Wochenschrift* 1967;92(43):1947-50. doi: 10.1055/s-0028-1106070
197. Shamley D. Pathophysiology: an essential text for the allied health professions: Butterworth-Heinemann 2005.
198. Balfour-Lynn I, Field D, Gringras P, et al. BTS guidelines for home oxygen in children. *Thorax* 2009;64(Suppl 2):ii1-ii26.
199. Teasdale G, Jennett B. Assessment of coma and impaired consciousness: a practical scale. *The Lancet* 1974;304(7872):81-84.
200. Burstein AH. Fracture classification systems: do they work and are they useful? *JBJS* 1993;75(12):1743-44.
201. Altman DG, Royston P. What do we mean by validating a prognostic model? *Statistics in Medicine* 2000;19(4):453-73.
202. Peek N, Arts DG, Bosman RJ, et al. External validation of prognostic models for critically ill patients required substantial sample sizes. *Journal of Clinical Epidemiology* 2007;60(5):491-501. doi: 10.1016/j.jclinepi.2006.08.011
203. Vogt WP, Johnson RB. Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences: A Nontechnical Guide for the Social Sciences: Sage 2011.
204. Efron B, Tibshirani RJ. An introduction to the bootstrap: CRC press 1994.
205. Liu VX, Fielding-Singh V, Greene JD, et al. The Timing of Early Antibiotics and Hospital Mortality in Sepsis. *American Journal of Respiratory and Critical Care Medicine* 2017;0(ja):null. doi: 10.1164/rccm.201609-1848OC
206. Seymour CW, Gesten F, Prescott HC, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *New England Journal of Medicine* 2017;376(23):2235-44. doi: 10.1056/NEJMoa1703058
207. Boardman SR, Chris; Robson Wayne; Daniels Ron. Prehospital management of a patient with severe sepsis. *Journal of Paramedic Practice* 2009;1(5):183-88.
208. Royal College of Physicians. National Early Warning Score (NEWS): Standardising the assessment of acuteillness severity in the NHS., 2012.
209. Corfield AR, Lees F, Zealley I, et al. Utility of a single early warning score in patients with sepsis in the emergency department. *Emergency Medicine Journal* 2014;31(6):482-7. doi: http://dx.doi.org/10.1136/emermed-2012-202186
210. Keep JW, Messmer AS, Sladden R, et al. National early warning score at Emergency Department triage may allow earlier identification of patients with severe sepsis and septic shock: a retrospective observational study. *Emergency Medicine Journal* 2015 doi: 10.1136/emermed-2014-204465

211. Cildir E, Bulut M, Akalin H, et al. Evaluation of the modified MEDS, MEWS score and Charlson comorbidity index in patients with community acquired sepsis in the emergency department. *Intern* 2013;8(3):255-60. doi: http://dx.doi.org/10.1007/s11739-012-0890-x
212. Jaimes F, Garces J, Cuervo J, et al. The systemic inflammatory response syndrome (SIRS) to identify infected patients in the emergency room. *Intensive Care Medicine* 2003;29(8):1368-71.
213. Slade E, Tamber PS, Vincent JL. The Surviving Sepsis Campaign: raising awareness to reduce mortality. *Critical Care* 2003;7(1):1-2.