Recognition of the deteriorating patient

Sarah Forster BSc (Hons) BMBS MRCP

Student ID: msxsf6

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

October 2023

Table of Contents

Table of Tables

Table of Figures

Abstract

Background

The discrimination between which hospital inpatients are stable, and which are at risk of clinical deterioration, has been the focus of significant research over the last 32 years since the concept of using vital sign derangement to alert a specialised team of people was first mooted. In the NHS, the National Early Warning Score has been mandated since 2017 across all adult patients outside obstetrics. However there have been concerns that it may not be equally predictive in all patient groups and may negatively impact hospital systems due to the demand generated by scores above set escalation thresholds.

Aim

To investigate the impact of NEWS2 on patients and hospital systems.

Methods

An initial literature review was performed in order to describe the current evidence base and define the research questions. A large outcomes-linked vital signs database was then analysed to determine the impact of introducing NEWS2 into a large teaching hospital with a mature electronic observations and task escalation system, before examining the predictive accuracy of NEWS2 in different population groups and investigating the possibility of improvements based on pattern of scoring. To complement this a qualitative study of the role of nursing concern in recognition and escalation of deteriorating patients was performed.

Results

The first study demonstrated an increase in demand following introduction of NEWS2, with a heterogeneity in accuracy of predicting risk of outcome of death within 24 hours between medical and surgical inpatients. This variation in prognostic ability was further demonstrated in a respiratory population and across cohorts defined by primary diagnosis and age. It was also demonstrated that improvements

in risk prediction could be made in all cohorts through the addition of simple pattern values, with maximum score in the preceding 24 hours providing the most additional information of the values. The qualitative study demonstrated that nursing staff employ several factors independently of NEWS2 when assessing a patient's clinical status and making a decision of whether to escalate for medical review.

Conclusions

This thesis has identified a variation in how different cohorts within a hospital population behave and the subsequent impact on predictive ability of NEWS2. The identification of pattern factors that could be incorporated into all systems, including those still using paper, is important as it could easily be integrated into future iterations. The clarification of the role of nurse concern in escalating patients at risk of deterioration should also be considered in future systems to improve risk prediction.

Structure of thesis

Question	**Chapter 2-** Effect of implementing the NEWS2 escalation protocol in a large acute NHS trust	**Chapter3-** Examining the performance of NEWS2 in patients with respiratory disease	**Chapter 4-** Analysis of impact of specialty and age on performance and the potential use of pattern to improve performance	**Chapter 5-** A Critical Decision Methods Study Of Nurse Concern In The Setting Of Early Warning Score Use.
Relationship between workload and patient deterioration	Analysis of workload, using escalations and NNE as a surrogate, across all areas	Analysis of differences in workload in patients with respiratory disease using COPD as paradigm	Analysis of differences in workload across diagnostic groups defined by ICD10	Exploration of perception of workload in relation to patient deterioration
Predicting which patients will go on to deteriorate	Regression analysis of factors predicting increased risk of mortality			Exploration of experience of NEWS2 in risk prediction
Evaluation of pattern as a possible predictor of risk		Analysis of Mean, minimum, maximum, standard deviation and change in respiratory patients	Analysis of Mean, minimum, maximum, standard deviation and change across diagnostic groups	Exploration of experience of how nursing staff use NEWS2 including trends
Additional features used to evaluate patient risk of deterioration	Identification that patients are escalated who don't meet the threshold set out in scoring escalation protocol			Analysis of factors used by nursing staff to evaluate patient risk at bedside

Abstracts arising from thesis

Academy of Medical Sciences Midlands Research Festival: Oral Presentation

Forster S; McKeever TM; Shaw DE. Identifying Patterns In Admissions, Mortality And Workload In Surgical And Medical Inpatients In An Acute Hospital Over A 4 Year Period From 2016-2019. AMS Midlands Research Festival 30th March 2022

BRC Joint Respiratory Research Day: Hybrid Poster Presentation

Forster S; McKeever TM; Shaw DE. Patterns In NEWS2 Scores As An Additional Tool In Predicting Outcome In Acute Respiratory Admissions. BRC Joint Research Day 11th November 2021

European Respiratory Society Annual Congress: Thematic Poster

Forster S; McKeever TM; Shaw DE. Analysis of NEWS2 trend to predict risk of death in COPD. ERS Annual Congress (virtual), 7th September 2020

Publications arising from thesis

Forster S; McKeever TM; Churpek M; Gonem S; Shaw DE. Predicting outcome in acute respiratory admissions using patterns of National Early Warning Scores. Clin Med (Lond) 2022 Vol. 22 Issue 5 Pages 409-415

Forster S; McKeever TM; Shaw DE. Effect of implementing the NEWS2 escalation protocol in a large acute NHS trust: a retrospective cohort analysis of mortality, workload and ability of early warning score to predict death within 24 hours. BMJ Open 2022 Vol. 12 Issue 11 Pages

Publications relating to this work prior to PhD period

S. Forster, G. Housley, T. M. McKeever and D. E. Shaw. Investigating the discriminative value of Early Warning Scores in patients with respiratory disease using a retrospective cohort analysis of admissions to Nottingham University Hospitals Trust over a 2-year period. BMJ Open 2018 Vol. 8 Issue 7

Other publications arising from this research period (not included in thesis)

S. Gonem, A. Taylor, G. Figueredo, S. Forster, P. Quinlan, J. M. Garibaldi, et al. Dynamic early warning scores for predicting clinical deterioration in patients with respiratory disease. Respir Res 2022 Vol. 23 Issue 1 Pages 203

Declaration of work performed personally

I was an NIHR academic clinical fellow during the period of August 2015 to August 2018. During this period initial work was performed examining the potential impact of introducing the National Early Warning Score into the respiratory population in Nottingham, with a particular focus on projected increase in workload.

During this period I developed the fellowship application which formed the basis of this PhD, with a Nottingham Hospitals Charity Fellowship allowing me to take time out of clinical work to perform the work.

Initial work involved the literature review which forms the first chapter and allowed the full definition of the research questions. Following this the protocol was written for, 'Improving understanding of patient deterioration through statistical analysis of previously collected, pseudo-anonymised clinical data linked to outcomes to profile differences between patients who go on to experience a serious adverse event versus those who don't', all of the relevant permissions from the Health Research Authority, the University and Nottingham University Hospitals Trust were sought.

During the period awaiting approvals, I undertook several courses within the University of Nottingham to develop the skills in statistics, data management and qualitative research methods which were required to successfully complete this project.

Following approval and creation of the database which formed the foundation of the data work in this project, analysis was performed under the supervision of Professors McKeever and Shaw. With Professor McKeever checking the statistical validity and accuracy of both methods and results.

Based on the initial findings of this work, and the literature review, the study protocol and ethics approval for, 'Identifying Triggers For Nurse Concern: A Critical Decision Methods Study' were written and obtained. I then recruited, took consent from and interviewed the 20 participants in the study. I then transcribed all of the interviews. This study was supervised by Professor Sharples who is a specialist in human factors research and advised on validity of analysis of transcripts.

COVID Impact Statement

The COVID pandemic impacted all work carried out in this period. In the case of this thesis it was affected in 3 ways. Firstly by delaying the work in general as I returned to clinical work as a senior intensive care registrar for several months during the first wave. Secondly, the impact of coronavirus on the patients admitted to the hospital meant that data from January 2020 onwards could not be included in the analysis as findings would not be valid in a hospital population outside of the pandemic. This removed 6 months of data which would otherwise have been included. Thirdly, the impact on nursing staff made recruitment for the qualitative study addressing nursing concern extremely challenging.

Acknowledgements

First of all I would like to thank the Nottingham Hospitals Charity, and particularly Barbara Cathcart and Dr Harish Vyas for seeing the potential in this project and for awarding me the William Colacicchi Fellowship which allowed me to take time out of clinical work to complete this work. Your support and interest has been so important.

The fact that this PhD has been completed at all is in large part due to the support of my three supervisors.

Firstly, Professor Dominick Shaw whose dissatisfaction with current systems and evidence for recognition of the deteriorating patient led us to start this work back in 2015. His encouragement, constructive criticism and wealth of experience have helped me develop the skills needed to complete the work detailed here. His foresight when the Nervecentre platform was developed is also the reason we have access to the hospital data in Nottingham.

The unsung hero of many projects is the statistician, but Professor Tricia McKeever has been the glue that has held this project together, particularly during the disruption caused by the COVID pandemic and the interruption of this project to return to clinical work. She stepped up to be lead supervisor and provided significant pastoral support while keeping me on the straight and narrow in terms of analysis and interpretation of data. For all of this I cannot thank her enough.

Special thanks also go to the supervisor of my qualitative work, Professor Sarah Sharples. She has guided on all of the human factors elements and provided a different viewpoint on the quantitative elements of the PhD. She has also provided copious advice on getting through life while undertaking a PhD in a pandemic.

Finally, to my family. My parents who have always encouraged me to follow the path less travelled in my winding career through respiratory, academia and intensive care. But most of all to my husband James. His support and encouragement to keep going has helped me finish, and marrying him will always be the highlight of my PhD.

Chapter 1 Literature Review

1.1 Introduction

Early warning systems are a clinical tool designed to support the recognition of and response to patient deterioration in a hospital setting. This review outlines the context of their development and the role of vital signs within these systems. It then goes on to highlight the measures which have been taken to further refine the process of identifying and responding to patients in a timely manner and discusses the evidence currently available for each step in that process. The final section considers the current challenges, and how this thesis contributes to the existing knowledge base.

It should be noted that although workload is referred to frequently throughout this thesis, this is by convention in line with previous research. What is actually measured here is a combination of workload, in terms of burden of increasing observation frequency, and demand in terms of number of escalations.

1.1.1 Context

In order to understand the reasons for developing early warning systems it is necessary to understand the context of patient management in hospital. There were 16.6 million completed hospital admission episodes recorded in England and Wales during the period April 2017 to March 2018 [1] These episodes fall into one of two categories. Planned or elective patients are admitted directly to their specialty area at a time when resources are available and with a plan for managing their admission based on predicted recovery trajectory. In contrast, unplanned or emergency admissions are triaged through the emergency department or via acute admissions areas. These areas have a proportionately higher level of medical and nurse staffing available to investigate and monitor patients during initial examination and management. From here patients are transferred, as appropriate, to a higher level of care, to theatres or to specialty-based wards for ongoing investigation and treatment. Complications are more common in emergency admissions, however, all patients have the potential for unexpected deterioration at any time.

During normal working hours, 0800-1600 Monday to Friday in surgical specialties and 0900-1700 in medicine, patients are managed by a ward-based team of doctors under the guidance of a consultant. Outside of these hours, and particularly overnight, there is a reduction in both medical and nursing staffing levels on wards, as well as allied health professional and administrative staff presence. This distinction is important as out of hours working accounts for 75% of the week.

Historically, continuity during the out of hours period was provided through junior doctors covering a weekend as a single shift as part of a firm structure, with nursing staff contacting their on-call doctor directly by pager or 'bleep'. Following implementation of the European Working Time Directive, a 48-hour working week meant that, with no significant increase in the workforce, only a small team of doctors was available to cover the hospital out of hours in order to remain compliant with safe-working guidelines. Two approaches to managing out of hours working have since been employed to manage the available resources. The first is dividing up the hospital by geographical location or specialty grouping and allocating each area a doctor for nursing staff to contact directly regarding clinical queries or tasks. The second, which has now been adopted by the majority of hospitals in the UK, is 'Hospital at Night' working, an initiative spearheaded by the Department of Health and piloted across 4 acute trusts from 2003 [2]. Hospital at Night allows out-of-hours tasks to be collated, reviewed and allocated by a central coordinator, typically a senior nurse, to a small multidisciplinary team including doctors, nurses and healthcare support workers. This has the advantage of filtering out non-urgent or inappropriate tasks without calling a team member away from their current activity. In addition, it allows tasks to be assigned to the person with the most appropriate skill set and ensure workload is balanced across the on-call team. As part of Hospital at Night working there is also a dedicated handover period at the start of each shift for important clinical information to be conveyed in verbal or electronic form.

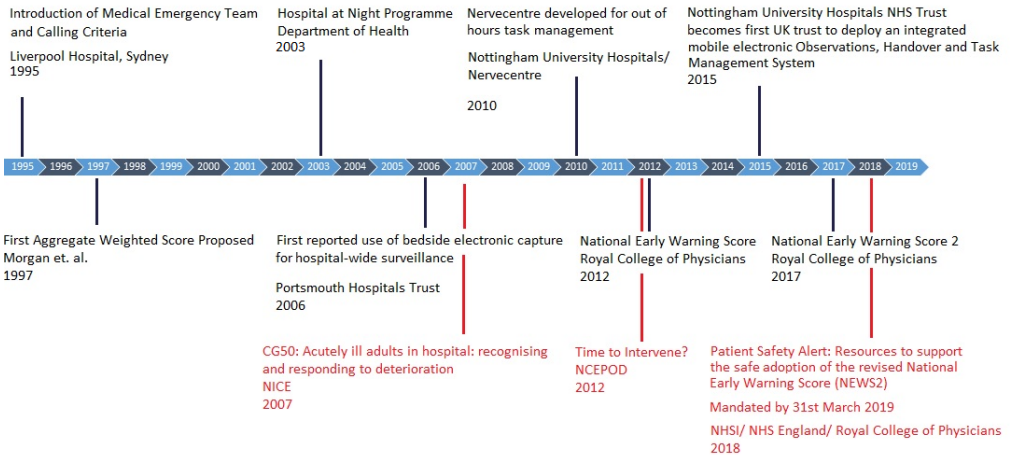
The number and seniority of doctors resident in the hospital varies during the out-of-hours period. A limited number of consultants are present during the day at weekends, with junior doctors available to act on plans from ward round and provide ongoing management in order facilitate clinical progress. At night, outside of

emergency admission, higher dependency and theatre areas, consultants are available for advice by phone, and registrars act as the most senior clinician resident to make patient and bed management decisions for medicine and surgery.

The presence of a smaller medical team out of hours places the onus on nursing staff to recognise when a patient needs further clinical review and communicate that need effectively. From admission to discharge, all patients have vital signs observations recorded by nursing staff or healthcare assistants. A set of vital signs observations is a combination of physiological measurements recorded to monitor the stability of the major organ systems, response to treatment and in order to recognise clinical deterioration. Airway and breathing are observed through recording of respiratory rate and oxygen saturations. Cardiac stability and fluid status are monitored through trends in heart rate and blood pressure. Possible presence of inflammation or infection uses temperature as a guide. Hydration and kidney function are monitored through measurement of absolute values of and trends in urine output, and general overview and neurological function is most commonly assessed using the AVPU scale (**A**lert, alert to **V**oice, alert to **P**ain, **U**nresponsive). All vital signs values documented by nursing staff have a normal range based on healthy individuals and our current understanding of what constitutes abnormal derangement.

Figure 1-1 outlines the progression of development over the last 30 years of early warning systems and highlights instances where patient safety reviews and introduction of new technologies have stimulated and guided progress

Figure 1- 1 Timeline of development of Early Warning Scores



1.2 Antecedents to serious adverse event

The outcomes most commonly used when monitoring inpatient progress are length of stay and serious adverse event. In terms of patient outcome a serious adverse event is defined as unplanned admission to intensive care, in-hospital cardiac arrest or death in hospital.

Death during admission or within 30 days of discharge during the period April 2017 to March 2018 was 3.3% according to national statistics produced by NHS England and NHS digital [3]. This observed crude mortality rate was significantly higher in patients who had an unplanned admission to intensive care or who suffered an in-hospital cardiac arrest. To quantify this, of the 175,700 admissions to general adult critical care units in England and Wales over the same period the mortality before discharge from hospital that admission was 19.7% [4], while the 16,000 cardiac arrests reported had a mortality before discharge of 78% [5]. Identifying hospital patients at risk of clinical deterioration is key to improving outcomes, on the basis that a timely intervention has the potential to reverse decline in a proportion of patients and reduce serious adverse events including unexpected admission to ICU, cardiac arrest and death in hospital.

Several studies have examined patient trajectories prior to these events in order to determine whether they can be predicted and to identify reliable antecedents. Table 1-1 describes the results of studies examining vital signs observations and actions in the period antecedent to serious adverse event, defined here as cardiac or respiratory arrest, unplanned admission to ICU or a compound outcome of all three. In these studies 59-84% of patients exhibited abnormal vital signs observations up to 8 hours before serious adverse event [6-14]; Those studies reviewed included a combination of respiratory rate, oxygen saturations, heart rate, blood pressure, temperature and neurological status. Two studies also identified an increased risk of deterioration associated with multiple deranged vital signs [7, 14]. This association between antecedent vital sign changes and subsequent clinical deterioration in the majority of patients studied [7, 9, 10, 13, 15] led to the widespread convention that all but the most stable of patients should have vital signs documented at least every 6 hours.

Table 1-1 Studies examining antecedent to serious adverse event

Author	Year	Title	Population	No. of patients	Study design	Findings
Schein et. al. [13]	1990	Clinical Antecedents to In-Hospital Cardiopulmonary Arrest	Patients identified as suffering cardiac arrest between Jul-October 1987 at Jackson Memorial Hospital Medical Center	64 59 with cardiopulmonary arrest 5 Respiratory arrest	Retrospective notes study	84% of patients showed evidence of documented clinical derangement in 8 hours before event Vital signs obtained mean 5 +/-1 hour prior to deterioration Mean value in last set of vital signs (Standard error of mean): HR 99 (+/- 3) RR 29 (+/-1) T 37.2 (+/- 0.17) SBP 118 (+/- 3) DBP 71 (+/-2)
Franklin et. al. [12]	1994	Developing strategies to prevent inhospital cardiac arrest: analysing responses of physicians and nurses in the hours before event	Patients suffering cardiac arrest at Cook County Hospital, Chicago over 20 month period 1990-1991	150	Notes review Within 48 hours of arrest Q1- had patient been in ICU that admission Q2- had deranged vital signs been documented within 6 hours of arrest	66% of patients showed evidence of documented clinical derangement in 6 hours before event. 91% mortality during admission following cardiac arrest Cardiac arrests with antecedents: 25%- abnormalities documented but not escalated 43% seen by doctor but not escalated to ICU 32% ICU triage error

McQuillan et. al. [11]	1998	Confidential inquiry into quality of care before admission to intensive care	A large district general hospital and a teaching hospital	100 admissions to ICU	Prospective confidential inquiry on the basis of structured interviews and questionnaires	54% of patients received suboptimal care prior to ICU admission; admission delayed in 35-64% Admission to intensive care considered: Avoidable in 4.5-1.7% Probably avoidable in 4-7.5% Possibly avoidable in 32.5-41.5% Note- Suboptimal care defined by consensus opinion not definition
Goldhill et. al. [10]	1999	Physiological values and procedures in the 24 hour before ICU admission from the ward	Royal London Hospital- admissions to ICU over 13-month period from May 1995	923 admissions- 76 patients met study criteria	Prospective observational study	34% of admissions to ICU followed CPR 47% of admissions had chronic health problems according to APACHEII criteria. Average APACHEII ICU score for assessing illness severity was 19 pre-admission to ICU (possible range of score 0-79) 75% of patients were on oxygen 6 hours pre admission to ICU. Significant worsening of RR but not HR in 24 hours pre-ICU.
Hillman et. al. [9]	2001	Antecedents to hospital deaths	3 similar sized hospitals in NSW, Australia over 6 month period 8th July to 31st December 1996	778 deaths reviewed	Prospective notes review of all patients aged 14 and over.	171 had cardiorespiratory arrest 160 deaths occurred on ICU- of these 49 were unplanned admissions from ward Of those deaths withno cardiac arrest call or admission to ICU: 447 deaths had antecedents. **Of these:** 125 (28%) had $\geq$1 deranged vital signs in 8 hours prior to death 85% had DNR orders Of the 66 without DNR orders, 33 (50%) had $\geq$1 deranged vital signs in 8 hours before death; 17 (26%) had same deranged vital signs present for 48 hours before death

Hillman et. al. [8]	2002	Duration of life-threatening antecedents prior to intensive care admission	3 similar sized hospitals in NSW, Australia over 6 month period 8th July to 31st December 1996	551 admissions to ICU	Prospective notes review of all patients aged 14 and over.	In 8 hours before admission, derangement was seen in: RR 12% HR 16% BP 36% Fall in GCS 8% Worry 7%
Buist et. al. [7]	2004	Association between clinically abnormal observations and subsequent in hospital mortality: a prospective study	Dandenong Hospital- a 320 bed university affiliated teaching hospital over 33 weeks between May-December 1999	6303 admissions to study areas	Prospective observational study	1598 abnormal bedside observation in 564 patients**Events seen** and **OR for mortality (95% CI)**RR<6 OR 13.7 (2.9-64) Tachypnoea OR 6.1 (3.6-10.6) Bradycardia OR 0.9 (0.5-1.8) Tachycardia >130 OR 2.3 (1.3-3.9) Hypertension OR 0.6 (0.3-1.2) Hypotension OR 1.8 (1.2-2.6) SaO2 OR 1.4 (0.9-2.0) P<0.05 for difference in probability of mortality in those with event versus those without event. 88% risk of mortality with ≥4or more abnormal observations 67% of observations spontaneously resolved 22% of observations resolved with treatment on ward 4% led to unplanned operation 2.4% led to unplanned admission to ICU

Kause et. al. [6]	2004	A comparison of Antecedents to Cardiac Arrests, Deaths and Emergency Intensive care Admissions in Australia and New Zealand and the United Kingdom- the ACADEMIA study	69 Hospitals in UK 19 Hospitals in Australia 2 Hospitals in New Zealander	638 patients with 1° events (cardiac arrest/ death/ admission to ICU)	Multi-centre, prospective, observational study of patients >16 suffering cardiac arrest, death, or unplanned admission to ICU	Primary events- Total 638: 308 deaths, 141 cardiac arrests, 189 unplanned ICU admission 168 deaths with antecedents (20 no DNR) 112 cardiac arrests with antecedents (96 no DNR) 103 ICU admissions with antecedents (93 no DNR)485 antecedents in 15min-24 hours before 1° event: RR<5 4% RR>36 11% HR<40 3% HR>140 9% SBP<90 31% Drop in GCS 24%
Husband et. al. [16]	2014	The epidemiology of respiratory arrests in a teaching hospital	The Austin Hospital- a 400 bed teaching hospital in Melbourne, Australia	79 patients with respiratory arrest (82 arrests)	Retrospective observational audit	In the 12 hours before respiratory arrest: 21% occurred during a MET call 24% fulfilled MET criteria without triggering response 13% showed progressive hypoxia 12% had RR>25 5% had low GCS 52% had no discernible antecedents 9% had incomplete or missing charts

Anderson et. al. [14]	2016	The prevalence and significance of abnormal vital signs prior to in-hospital cardiac arrest	300 hospitals in the USA- The Get With the Guidelines resuscitation registry	7851	Post Hoc analysis of prospectively collected data	Looked at vital signs 1-4 hours before cardiac arrest Definition of abnormal: **<10 RR >20; <60 HR >100; <90 SBP** Definition of severely abnormal**: <8 RR>30; <50 HR >130; <80 SBP** **Abnormal (%)** / **Severely abnormal (%)** 0: 3189 (40.6) / 6802 (86.6) 1: 3043 (38.8) / 946 (12.1) 2: 1446 (18.4) / 96 (1.2) 3: 173 (2.2) / 7 (0.1) Stepwise increase OR of 1.53 for each additional abnormal vital sign or 1.62 for each severely abnormal vital sign documented

Abbreviations: HR, Heart rate (beats per minute); RR, Respiratory Rate (breaths per minute); T, Temperature (℃); SBP, Systolic Blood Pressure (mmHg); DBP, Diastolic Blood Pressure (mmHg); SaO2, oxygen saturations; GCS, Glasgow coma scale; MET, Medical Emergency Team

1.3 Monitoring to predict deterioration

The identification of vital sign derangement in the majority of patients prior to serious adverse events provides a window where it may be possible to alter trajectory. In order to take advantage of this window there needs to be effective monitoring in order to identify, and act on, these changes at the earliest possible stage.

Monitoring was defined by an international consensus conference in 2010 as, 'the assessment of a patient at predetermined intervals with the intention of 1) detecting abnormalities and 2) triggering a response if an abnormality is detected' [17]. Consequently, a monitoring system needs an afferent limb to detect evidence of deterioration, and an efferent limb to respond to it. Such a system can be broken down into four elements, surveillance and recognition comprise the afferent limb, and referral and response form the efferent limb.

1.3.1 Afferent limb- Surveillance

Surveillance requires the repeated collection of a core set of vital signs observations. Rigor of surveillance depends on accuracy and efficiency of data collection, documentation and data sharing. It is therefore important to consider these as individual components in order to determine how each could potentially be optimised.

1.3.1.1 Documentation: Chart design

Vital signs observation charts are a longstanding and ubiquitous tool in the hospital setting. They traditionally include the patient's respiratory rate, oxygen saturations, heart rate, blood pressure, temperature and some form of neurological assessment, recorded on paper and kept at the patient's bedside or at a nursing station. In their most basic form they are recorded at intervals, conventionally once per shift, but determined by ward practice or nurse concern. In this approach reliance is placed on nursing and medical staff to recognise and act on signs of deterioration, through taking a set of observations, escalating to the medical team or both.

There are two key features of vital sign documentation. These are frequency and accuracy of recording. Although there are no studies which explore the optimal frequency of observations in a deteriorating patient, there is a general agreement that the greater the perceived acuity the more frequently observations should be performed. This ranges from 12 hourly in the most stable patients

to every 15 minutes to monitor response to intervention where significant instability is felt to be present, guided by local protocols.

Many studies examining both antecedents and accuracy of vital sign documentation noted that a proportion of patients had no vital signs documented in a 24 hour period. Table 1-2 details the findings of studies which examined the accuracy and completeness of vital sign documentation before and after an intervention, whether that be a change in chart design, in documentation method, or addition of a Medical Emergency Team (MET). A clear lack of accuracy in documentation was noted at baseline across all studies. Respiratory rate was the most frequently neglected vital sign, missing in 52.2-75.3% [18-20] in unselected pre-intervention cohorts. The majority of interventions reported some degree of improvement following the associated intervention.

Strategies to improve accuracy of charting, and compliance with local protocols have focussed on chart design and education. Features such as graphical representation of values, introduction of normal ranges with colour coding to highlight derangement, or an associated track and trigger [21, 22] have been explored. A track and trigger score applies a score to each vital sign from 0 to 3 depending on how far outside the normal range it is. A single parameter track and trigger score uses these values individually against an escalation protocol. An aggregate weighted track and trigger score, also known as an early warning score, adds the individual vital signs scores together, with a protocol for clinical indicated at each score or scoring band.

These interventions, along with staff education, have demonstrated significant improvements in accuracy of recording and compliance with protocol [18, 20, 23, 24]. A proportion of this change can likely be attributed to a combination of staff engagement, education and the Hawthorne effect [25]. Nevertheless, persisting improvement demonstrated at one year after the introduction of an EWS [18]suggests that the need to calculate a score and act on it may contribute to a more comprehensive completion of vital signs recording in every observation set. As all components need to be recorded to generate a score, which can then be used to provide context to concerns and guidance regarding when the timing of the next observation set and appropriate escalation of the patient.

Table 1-2 Vital sign accuracy and the impact of associated interventions

Author	Year	Title	Population	No. of patients	Study design	Intervention	Findings
Arora et. Al. [23]	2005	Evaluation of CoViSTA – an Automated Vital Sign Documentation System – in an Inpatient Hospital Setting	Inpatients at Baltimore Veterans Affairs Medical Centre	60	Prospective randomised crossover trial	Introduction of Co-VISTA system for automatic bedside vital sign capture	Errors in data entry pre: 7/30 (23) Omitted 2/30 (7%) Errors in intervention group- 0/30 No Omissions
Mcbride et. Al. [18]	2005	Long-term effect of introducing an early warning score on respiratory rate charting on general wards	Portsmouth Hospitals NHS Trust: 6 wards: 2 x orthopaedic 2 x medical 2 x surgical	Stage1: 1251 Stage2: 1234 Stage3: 600	Prospective stepped wedge intervention of respiratory rate monitoring before intervention (week 0-17), immediately following intervention (weeks 23-30) and after a settling period (weeks 67-69)	introduction of new observation chart and modified early warning score in stepped manner	**Average % of occupied beds with at least one resp rate recorded in 24 hours**: Stage1: 29.5±13.5% Stage 2: 68.9±20.9% Stage 3: 91.2±5.6% Statistically significant increase in recording between stage 1-stage 3 (Fischer's exact p<0.001)
Gearing et. Al. [26]	2006	Enhancing patient safety through electronic medical record documentation of vital signs	University College Hospital, Tampa Bay: 20 bed cardiac step down 27 bed medical surgical unit	Cardiac step down: 613 Medical surgical: 623	Audit of vital sign completeness in paper versus electronic medical record documentation of vital signs	Examination of 2 wards with different data entry within same hospital (no new intervention)	**Errors**: Paper charting on medical/surgical unit: 157/613 (25.6%) Electronic medical record on cardiac step-down: 93/623 (14.9%)

Prytherch et. Al. [27]	*2006*	*Calculating early warning scores——A classroom comparison of pen and paper and hand-held computer methods*	*Sim setting- 84 fictitious observation sets*	*84 sets of observations (sim setting)*	*Prospective randomised crossover trial*	*Introduction of VitalPAC for input of vital signs observations.*	*Errors of input:* *Paper charting:* *37/504 data points (7.3%); 24/84 sets (27.4%)* *VitalPAC:* *12/504 data points (2.4%); 8/84 sets (9.5%)*
Chen et. Al. 2011 [19]	2008	The impact of introducing a medical emergency team on documentation of vital signs	23 Hospitals in Australia-cluster randomised as part of MERIT study	Control: 2357 MET: 3625	Cluster randomised control trial	Introduction of Medical emergency team	**Data sets with one or more values missing** **Control hospitals:** Baseline: 25% of 460 Implementation: 21% of 796 Study: 22% of 1101 **MET hospitals** Baseline: Missing 29% of 435 Implementation: Missing 26% of 899 Study: Missing 26% of 2291 No statistically significant difference
Cahill et. Al. [20]	2011	Introduction of a new observation chart and education programme is associated with higher rates of vital-sign ascertainment in hospital wards	Royal Prince Alfred Hospital-a University-affiliated teaching hospital in Sydney	Pre-intervention:104 2 weeks post: 147 3 months post: 119	Prospective before and after intervention study	Introduction of new observation chart including track and trigger protocol and education	**Completeness**: **Pre**: Full set: 47.6% Resp rate: 47.8% **2 weeks post:** Full set: 96.3% Resp rate: 97.8% **3 months post:** Full set: 96.4% Resp rate: 98.5% **Improvement in resp rate**: p<0.001 **Improvement in full set**: p <0.001

<table>
<tr>
<td>Elliott et. Al. [24]</td>
<td>2017</td>
<td>User compliance with documenting on a track and trigger-based observation and response chart: a two-phase multi-site study</td>
<td>10 health sites in Australia- 2-6 adult medical-surgical wards at each site.</td>
<td>Phase 1: 818
Phase 2: 1058</td>
<td>Two phase audit- Retrospective post-intervention Feb 2011, prospective post-intervention Feb 2012</td>
<td>Introduction of one of 5 new chart templates to act with pre-existing RRT

Phase 2 completed with 3 charts- ADDS; R4, R2- all with track and trigger elements</td>
<td>Error rates not documented for phase 1:
Phase 2: % Completion of vital signs
ADDS:
<table>
<tr><th></th><th>RR</th><th>Sats</th><th>HR</th><th>BP</th><th>T</th></tr>
<tr><td>Retro</td><td>90</td><td>95</td><td>96</td><td>96</td><td>92</td></tr>
<tr><td>Pro</td><td>97</td><td>97</td><td>99</td><td>100</td><td>97</td></tr>
</table>
R4:
<table>
<tr><th></th><th>RR</th><th>Sats</th><th>HR</th><th>BP</th><th>T</th></tr>
<tr><td>Retro</td><td>72</td><td>94</td><td>97</td><td>97</td><td>86</td></tr>
<tr><td>Pro</td><td>96</td><td>96</td><td>97</td><td>100</td><td>95</td></tr>
</table>
R2:
<table>
<tr><th></th><th>RR</th><th>Sats</th><th>HR</th><th>BP</th><th>T</th></tr>
<tr><td>Retro</td><td>88</td><td>90</td><td>91</td><td>93</td><td>86</td></tr>
<tr><td>Pro</td><td>94</td><td>94</td><td>94</td><td>95</td><td>89</td></tr>
</table></td>
</tr>
</table>

Abbreviations: HR, Heart rate (beats per minute); RR, Respiratory Rate (breaths per minute); T, Temperature (°C); SBP, Systolic Blood Pressure (mmHg); DBP, Diastolic Blood Pressure (mmHg); Sats, oxygen saturations; GCS, Glasgow coma scale

1.3.1.2 Documentation: Electronic charting

Electronic vital sign recording systems started appearing in the late 1990s. Earlier systems were criticised by clinical staff for the inability to directly record and view results at the bedside, leading to anecdotes of observations for an entire bay being temporarily recorded on paper towels. However, the majority of electronic observations platforms now utilise portable devices for direct data input, facilitating greater efficiency and a reduction in the opportunity for transcription error [26].

In addition, as the software now automatically calculates scores generated by track and trigger systems, complexity in calculation is masked at the clinical interface, allowing more intricate scores to be generated from vital signs observations without additional human error from miscalculation[27]. Other benefits include the ability to monitor patients remotely, and for more than one person to view a patient's observations at any one time. In the advent of hospital at night working out of hours, this allows clinical decision makers to have more information before prioritising review.

1.3.1.3 Night-time working as a paradigm for barriers to recording observations

Compliance with protocols for recording vital signs is more than a matter of optimising design, platform and staff education. The quantitative studies in table 1-3 demonstrate a clear deviation from set protocols relating to frequency of vital sign recording overnight. This is illustrated by both an overall reduction in observations measured overnight, and a significantly longer mean time until the next set of vital signs were documented when compared with patients in the same scoring band during the day. There were also noted to be a large number of missing or incorrect observations when compared with studies examining accuracy of vital sign documentation and adherence to timings throughout the day.

There appear to be two main factors involved in lack of adherence to charting protocols overnight. The first is resources. Several studies describe inadequate staffing levels as being a factor in completing observations in adherence to protocol [28-30]. In one study, while 85% felt scheduled observations were very important overnight, only 46% agreed there were enough staff at night to perform them on time and 48% felt the skill mix was inappropriate for the workload [30].

The second major factor is the frequently stated aim of creating a block of interruption free time to promote sleep and recovery [29-31]. This is attributed to the perception that rest is of greater value to the trajectory of a patient they perceive to be stable than rigidly following a protocol [32]. This

was variously done by timing observations around other interventions to deliver a gap in interference, or through missing observations where the patient, or patients in the bay, were asleep. In one study, 34% of responders stated they would only wake patients to take observations overnight if they were worried. In another paper 48% of night staff stated that they would omit taking vital signs observations overnight if requested by the patient [30], preferring to rely on an end of bed assessment and clinical judgement rather than observations. There were also reports of senior nurses encouraging colleagues not to wake patients [31].

Table 1-3 use of early warning scores at night

Author	Year	Title	No. of participants	Study design	Setting	Findings
Quantitative Studies						
Gordon C, Beckett D [33]	2011	Significant deficiencies in the overnight use of a standardised early warning scoring system in a teaching hospital	Ward: 121 Combined assessment unit: 8	Prospective observational study- patients scoring a SEWS ≥4 or with nursing concern overnight	Royal Infirmary, Edinburgh	No chart had urine output documented Ward / CAU SEWS total missing 67/121 (55%) / 4/8 (50%) Observations missing 77/121 (64%) / 3/8 (38%) SEWS total incorrect 26/121 (21%) / 2/8 (25%) Observation recorded wrong box 18/121 (15%) / 2/8 (25%)
Hands et. al. [34]	2013	Patterns in the recording of vital signs and early warning scores: compliance with a clinical escalation protocol	950 043 vital signs sets	Retrospective audit of vital signs documentation in comparison to hospital protocol	Queen Alexandra Hospital, Portsmouth Hospitals NHS trust	Time between last ViEWS recorded 08.00-11.59 and 20.00-23.59 and time to next observation (TTNO): **VIEWS band** / **Total obs 0800-11.59** / **Mean TTNO** / **Total obs 20.00-23.59** / **Mean TTNO** **0-1** 49468 6.46 88742 8.95 **2** 17109 6.07 29029 8.37 **3-6** 25276 5.64 38521 7.88 **7-8** 2281 4.91 27924 6.59 **>9** 951 4.22 964 5.17 P value for difference in TTNO <0.001 for all VIEWS bands 13% of total daily obs. measured between 23.59-05.59 (29% of 24 hour day)

Yiu et. al. [35]	2014	Into the night: factors affecting response to abnormal Early Warning Scores out-of-hours and implications for service improvement	109 patients	Service evaluation with prospective record review	Ysbyty Gwynedd, Bangor 210 medical beds	109 patients scoring NEWS ≥6 NEWS 6-8 escalated in 14/91 NEWS ≥9 escalated in 4/18- NB. None escalated to SpR despite protocol

	Number (%)	**Escalated (%)**
Persistent NEWS ≥6	47/109 (43)	8/47 (17)
DNAR	44/109 (40)	5/44 (11)
Frailty scale ≥5	82/109 (75)	16/82 (20)
COPD	71/109 (65)	8/71 (11%)

Qualitative Studies

Petersen et. al. [29]	2017	Barriers and facilitating factors related to use of early warning score among acute care nurses: a qualitative study.	18 nurses- 7 surgical 11 medical	Focus groups- 1) What are the barriers and facilitating factors in relation to adhering to monitoring frequency? 2) What are the barriers and facilitating factors in relation to informing doctors at EWS ≥ 3? 3) What are the barriers and facilitating factors in relation to initiating MET calls?	700 bed hospital in the capital region of Copenhagen, Denmark.	Lack of adherence to monitoring frequencies during busy periods and at night. Nurses expressed concern about patients' sleep and reluctant to disturb patients at night. Lack of resources identified as a key barrier to adequate monitoring, while increased staffing led to more efficient monitoring. EWS protocol not considered mandatory. Suggestions regarding automated monitoring or limiting to higher risk groups to lighten workload and facilitate better adherence. Nurses from surgical ward had lower threshold for requesting clinical review.
Hope et. al. [31]	2018	A fundamental conflict of care: Nurses' accounts of balancing patients' sleep with taking vital signs observations at night	44 staff for adherence levels audit 17 staff for interview phase- 9 face to face 8 telephone	Qualitative Interpretative- Semi-structured interviews exploring decisions surrounding use of early warning score at night	Queen Alexandra Hospital, Portsmouth Hospitals NHS trust	All identified sleep as a core part of night time care. Seen as important to recover, staff felt interruptions would be detrimental. Vital signs also identified as core tool for detecting deterioration. One staff member had never woken someone at night for obs. Reported a nurse in charge had said not to wake patients. Several reported using clinical judgement in relation to patient stability effects caused by medications and underlying chronic disease.

Recio-Saucedo et. al. [30]	2018	Relationships between healthcare staff characteristics and the conduct of vital signs observations at night: Results of a survey and factor analysis	497 staff working at least one night shift per year	Exploratory descriptive study using online survey	Queen Alexandra Hospital, Portsmouth Hospitals NHS trust	54% felt taking obs. at night very disruptive 85% felt schedule obs. very important overnight. 48% would omit obs. if requested by patient. 34% would only wake up patient if worried 46% felt there were enough staff at night to perform scheduled observations on time 48% felt skillmix inappropriate for workload 71% agreed all patients with EWS $\geq$6 are escalated to H@N for review
Foley C, Dowling M [28]	2019	How do nurses use the early warning score in their practice? A case study from an acute medical unit	8 nurses were interviewed	Descriptive case study design using semi-structured interviews, observation and document analysis.	Acute medical short stay ward with 15 beds in large regional hospital in Ireland	Clinical experience ranging 5-22 years Task driven approach to EWS management- low priority and not recorded based on previous score. Inadequate staffing noted during study. Nurses rarely described EWS as method for patient assessment

Abbreviations: SEWS, standardised early warning score; NEWS, National early warning score; COPD, chronic obstructive pulmonary disease; DNAR, Do not attempt resuscitation order; obs, vital sign observations set.

1.3.1.4 Wearable Monitoring

The reduction in both size and cost of digital technology has led to the development of wearable devices that can be used for intermittent or continuous vital sign monitoring. This removes the need for staff to physically take observations and potentially reduce patient disturbance, particularly in a bay setting. This may also have beneficial infection control implications. A further significant consideration is the removal of transcription error in recording vital signs through communication of values wirelessly to scoring platforms. Although studies have shown that the majority of patients would be prepared to wear such a device [36], some concern was raised regarding the reduction in nurse-patient contact that this could create [37], both from a caring and assessment perspective. The role of the end of the bed risk assessment provided by nursing staff is addressed in a later section exploring the role of nurse concern in identifying patients at risk of deterioration

Several different designs of wearable monitor are currently in development, with wrist, ear and patch-based solutions present in the literature. However, there is no evidence regarding impact on management of patients outside of a high dependency setting due to inadequate transmission of accurate data and generation of artefacts. In two recent trials of devices, one reported missing data of 8.4-10.1% due to failures in connection and data storage [38], while the second reported only 34% data completeness for respiratory rate monitoring [39]. Further studies are underway [40], with ongoing progress in the technology suggesting it is likely to be only a matter of time before wearable monitors acquire an accuracy that will allow meaningful evaluation in a general ward setting.

1.3.2 Afferent Limb- Recognition

In order for accurate surveillance of vital signs to impact patient trajectory there needs to be a mechanism to reliably recognise when vital signs values indicate a potential problem and activate a clinical response via the efferent limb. Normal ranges for each vital sign indicate the limits of the values expected in the absence of pathology, acute or chronic, with the cut offs providing trigger points. These ranges have been developed through a combination of clinical experience and statistical analysis and form a core component of all systems that utilise physiological criteria to identify at risk patients, either independently or as part of a more complex tool. Two distinct pathways for applying these ranges have developed from this concept. These are the single parameter track and trigger system and the aggregate weighted early warning score.

1.3.2.1 Single parameter track and trigger scores

The single parameter track and trigger system applies upper and lower limits of normality to each vital sign. If a single vital sign is outside of these parameters then a response is triggered. The first use of this concept was in Australia where it was developed into a set of physiological calling criteria (Figure 1-2) for the Medical Emergency Team [41], first introduced to New South Wales in 1990.

Figure 1- 2 Medical emergency team physiological calling criteria [39]

Abnormal Physiology	
Temperature (°C)	< 35.5 or >39.5
Systolic blood pressure (mmHg)	<100 or >200
Respirations/minute	<10 or >30
Pulse rate/ minute	<40 or >120
Urine output over 24 hours (ml)	<500
Decreased or altered level of consciousness	

1.3.2.2 Aggregate weighted early warning scores

Patients with multiple deranged vital signs observations are more likely to have a serious adverse event [14] and have a higher mortality rate [7]. This information was used to develop aggregate weighted scores, with the first such example appearing in a conference abstract in 1997 [42]. These apply a weighting to each vital sign dependent on where the value sits in relation to the proscribed normal range. The numbers assigned to each vital sign according to this weighting are then added up to give a score, which in turn is applied to a protocol specifying clinical response. The original score used five physiological parameters of systolic blood pressure, heart rate, respiratory rate, temperature and level of consciousness on the AVPU scale. Subsequent scores have added oxygen saturations, while others have also included oxygen delivery and urine output.

1.3.2.3 The evidence base for single parameter and aggregate weighted scores in an unselected population

Single parameter and aggregate weighted scores were initially developed through analysis of small volumes of data combined with clinical expertise [42]. Electronic observations systems have accumulated large clinical databases linked to outcome that can be retrospectively analysed to design and statistically validate more complex scores. Table 1-4 describes studies that have retrospectively evaluated the statistical discrimination of different early warning scores in an unselected population, i.e. a population not specifically limited by disease, specialty or age-based clustering. This use of such large data sets allows rare outcomes, such as ICU admissions and cardiac arrest, to be adequately powered for significance in terms of statistical ability to predict outcome. The studies outlined in table 1-4 demonstrate that aggregate weighted scores have a superior statistical discrimination in predicting a serious adverse event when applied retrospectively to large data sets.

The Royal College of Physicians developed the National Early Warning Score (NEWS) as a step towards standardising recognition of the deteriorating patient. This would mean staff moving to new hospitals would be familiar with the system and potentially allow for more direct comparison between hospitals. Statistically, NEWS version 1 compares favourably with other scores developed, with an area under the curve for predicting ICU admission within 24 hours of 0.65-0.86 and for predicting any serious adverse event of 0.66-0.87. The range of accuracy is due to variability in statistical discrimination based on the population the score was applied to. All scores which have been applied to different populations in multiple centres, including NEWS, ViEWS (the Vitalpac TM Early Warning Score) and eCART (Electronic Cardiac Arrest Risk Triage), display this variation when applied to different populations. In each case, the best statistical performance is seen in the populations in which they were developed. This is likely due to an element of statistical over-fit based on characteristics of the local population and healthcare systems.

The statistical discernment of a score in predicting an outcome is also reliant on the time lag between the vital signs being collected and the outcome being observed. Early studies, and those without access to a more sophisticated data set, are reliant on admission observations as a basis for predicting outcome at any point during that admission [43-49]. However, the use of electronic observation platforms to collect data throughout admission has generated databases that enable a more clinically relevant analysis of deterioration within 24 hours [44, 50-54], and the possibility of analysing vital sign trend immediately prior to outcome[55, 56].

Table 1-4 Statistical evaluation of Early Warning Scores

Author	Year	Title	No. of patients	Study design	Population	Score used and outcomes	Findings
Subbe et.al. [49]	2001	Validation of a modified Early Warning Score in medical admissions	709	Prospective observational study Single centre	56 bedded medical admissions unit of a district general hospital	Modified early warning score Outcomes: HDU/ICU admission; attendance of arrest team for cardiorespiratory emergency; Death at 60 days	For scores on admission- combined endpoints had area under ROC curve for: MEWS: 0.67 MEWS + age: 0.72 Maximum score of ≥5 associated with: **Risk of:** **OR** **95%CI** Death 5.4 2.8-10.7 ICU 10.9 2.2-55.6 HDU 3.3 1.3-9.2
Duckitt et. al. [48]	2007	Worthing physiological scoring system: derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based, single centre study.	Derivation set: 3184 Validation set: 1102 Combined set for comparison stats: 4286	Prospective observational study. 2-phases of data collection. Databases used for derivation and validation of Worthing score Single centre	Emergency admissions unit of 602 bed district general hospital.	Original Early Warning Score vs Worthing Early Warning Score 1° outcome: In hospital mortality Scores compared on combined derivation and validation set. **Note**: Worthing score retrospectively applied.	Score analysis on combined data set **Score used** **AUC** **95% CI** **P for fit** Worthing score 0.72 0.66-0.79 0.565 Worthing + age 0.81 0.65-0.71 Score analysis validation data set: **Score used** **AUC** **95% CI** **p** Worthing score 0.74 0.71-0.77 <0.001 Original EWS 0.68 0.65-0.71 <0.001

Smith et al. [47]	2008	Review and performance evaluation of aggregate weighted 'track and trigger' systems	9987 patients- 1x admission vital sign set from each	Retrospective application of 31 scores to previously collected vital signs	58 bedded medical admissions unit May-December 2006	Vital signs at admission linked to outcome following completed admission episode 1° outcome: in hospital mortality	Acute medical admissions- top 4 performing scores: **Score** **AUC** **95%CI** **P** **Bakir EWS 2005** 0.782 0.767-0.797 <0.001 **Subbe MEWS 2001** 0.761 0.745-0.777 <0.001 **Subbe MEWS 2007** 0.728 0.711-0.745 <0.001 **Subbe MEWS 2002** 0.722 0.705-0.740 <0.001
Cuthberts on et al. [46]	2010	The use of combined physiological parameters in the early recognition of the acute medical patient	Medical- 300 (respiratory cohort was also studied)	Prospective cohort study- vital signs collected from firs 48 hours of admission Single centre	Acute medical admissions Aberdeen royal infirmary- 900-bed acute hospital	1° outcome: ICU admission All retrospectively applied to prospectively identified cohorts.	Acute medical admissions- **Score** **Sens** **Spec** **AUC** **PPV%** **NPV%** **EWS** 0.83 0.70 0.81 91.4 53.1 **PART** 0.79 0.77 0.84 92.8 49.0 **MEWS** 0.83 0.79 0.87 93.6 55.8 **SEWS** 0.95 0.77 0.95 94.0 79.7

Prytherch et. al. [45]	2010	ViEWS- Towards a national early warning score for detecting adult inpatient deterioration	35585 consecutive, completed acute medical admissions	Retrospective cohort study	Acute medical admissions unit of Portsmouth Hospitals NHS trust	Applied 33 aggregate weighted scores to data. 1° outcome: In hospital mortality within 24 hours of observation set.	5 top scoring aggregate weighted scores: Note- ViEWS derived and validated on Portsmouth data set. **Score** / **AUROC** / **95%CI** / **P value** **ViEWS** 0.888 0.880-0.895 <0.001 **Subbe** 0.850 0.841-0.859 <0.001 **Duckitt** 0.849 0.839-0.858 <0.001 **Paterson** 0.843 0.833-0.853 <0.001 **Barlow** 0.842 0.833-0.852
Smith et. al. [57]	2013	The ability of the National Early Warning Score to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission and death	35585 consecutive, completed acute medical admissions	Retrospective cohort study	Acute medical admissions unit of Portsmouth Hospitals NHS trust	Death, cardiac arrest and unanticipated ICU admission within 24 hours of a vital signs data set	AUROC for NEWS- outcomes within 24 hours of an observation set: **AUC** / **95% CI** **Cardiac arrest** 0.722 0.685-0.7 **ICU admission** 0.857 0.847-0.8 **Death** 0.894 0.887-0.9 **Composite** 0.873 0.866-0.8

Churpek et. al. [44]	2014	Using electronic health record data to develop and validate a prediction model for adverse outcomes on the wards	56649 controls 109 cardiac arrest patients 2543 ICU transfer patients	Retrospective cohort study	All admissions to academic medical centre in USA with 500 inpatient beds	Outcomes: Cardiac arrest ICU transfer	**Outcome**	**Cardiac arrest model**	**ICU transfer model**	**ViEWS**
							Ever experienced event			
							Cardiac arrest	**0.88** (0.84-0.91)	-	**0.78** (0.73-0.83)
							ICU transfer	-	**0.77** (0.76-0.78)	**0.73** (0.72-0.74)
							Experienced event within 24 hours			
							Cardiac arrest	**0.88** (0.88-0.89)	-	**0.74** (0.72-0.75)
							ICU transfer	-	**0.76** (0.76-0.76)	**0.73** (0.72-0.73)
Kovacs et. al. [51]	2016	Comparison of the National Early Warning Score in non-elective medical and surgical patients	Medicine: 48747 Surgery: 20626	Retrospective database study	Admissions to Portsmouth Hospitals NHS Trust May 2011-December 2013	Outcomes: Death, cardiac arrest and ICU admission within 24 hours	**All observations:**	**Surgical**	**Medical**	
							Death	**0.914** (0.907-0.922)	**0.902** (0.898-0.905)	
							Cardiac Arrest	**0.765** (0.733-0.792)	**0.747** (0.735-0.759)	
							Unplanned ICU	**0.860** (0.853-0.868)	**0.864** (0.857-0.870)	
							Combined	**0.874** (0.868-0.880)	**0.874** (0.871-0.877)	
							Random observations			
							Death	**0.919** (0.892-0.944)	**0.929** (0.920-0.937)	
							Cardiac Arrest	**0.722** (0.661-0.779)	**0.744** (0.720-0.767)	
							Unanticipated ICU	**0.831** (0.810-0.851)	**0.871** (0.856-0.886)	
							Combined	**0.848** (0.832-0.864)	**0.888** (0.880-0.895)	

Hodgson et. al. [58]	2017	Validation of the National Early Warning Score to predict outcome in patients with COPD exacerbation	Acute medical patients 20 415 COPD patients 942	Observational cohort study	Admissions to 2 acute medical units- Worthing Hospital and St Richards Hospital	Outcome: Inpatient mortality NEWS automatically calculated (EWS in current use)	AUC for inpatient mortality based on admission NEWS: **0.75 95% CI 0.74-0.76** Note: All other values in this paper relate to a COPD rather than general admissions cohort.
Ghosh et. al. [52]	2018	Early Deterioration Indicator: Data-driven approach to detecting deterioration in general ward	2097	Retrospective cohort study	Admissions to a community hospital in Phoenix, Arizona	Deterioration in 24 hours- defined as death or transfer to higher level of care.	**AUC deterioration** / **AUC mortality** **EDI** 0.866 / 0.70 **NEWS** 0.657 / 0.63 **MEWS** 0.649 / 0.57
Watkinson et. al. [53]	2018	Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data	53395	Retrospective cohort study	Admissions to four Oxford hospitals 1st October 2015- May 2017	Outcome: Composite of unanticipated admission to ICU, cardiac arrest or death within 24 hours of an observation set	Area under the curve based on prediction of composite outcome: **Score** / **AUC** / **(95%CI)** **Manual CEWS** 0.868 (0.864-0.872) **NEWS** 0.867 (0.863-0.871) **MEWS (Subbe '01)** 0.821 (0.817-0.825) **CART** 0.729 (0.725-0.734)

Lagadec et. al. [59]	2019	The efficacy of twelve early warning systems for potential use in regional medical facilities in Queensland, Australia	331- 159 Index 172 Control Matched for demographics, LOS, comorbidities	Retrospective case control study	2 private regional hospitals in central Queensland	Outcomes: Specificity, sensitivity and AUROC in predicting serious adverse event. Time before event scores would trigger a medical emergency team review	Area under curve for predicting SAE- 5 top performing scores: **EWS** **AUROC** **95%CI** **Compass** 0.747 0.73-0.76 **NEWS** 0.741 0.73-0.75 **Q-ADDS** 0.723 0.71-0.74 **MADDS** 0.705 0.69-0.72 **ADDS** 0.701 0.69-0.71
Pimental et. al. [54]	2019	A comparison of the ability of the National Early Warning Score and the National Early Warning Score 2 to identify patients at risk of in-hospital mortality- a multi-centre database study	251 266 48898 identified to be at risk of T2RF 1394 documented T2RF	Retrospective database study	Admissions to Oxford Unviersity Hospitals and Portsmouth University Hospitals	Outcomes within 24 hours: Unanticipated ICU admission Cardiac arrest Composite outcome	**T2RF** **At risk of T2RF** **Not at risk of T2RF** **ICU admission** **NEWS** **0.806** (0.786-0.826) **0.814** (0.808-0.821) **0.841** (0.837-0.845) **NEWS2** **0.816** (0.796-0.836) **0.815** (0.808-0.821) **0.833** (0.829-0.837) **Cardiac Arrest** **NEWS** **0.701** (0.654-0.749) **0.756** (0.744-0.769) **0.785** (0.776-0.794) **NEWS2** **0.706** (0.658-0.753) **0.741** (0.728-0.754) **0.768** (0.760-0.777)

Pimental et. al. [54] cont.							**Composite outcome**
							NEWS **0.835** (0.824-0.847) **0.858** (0.855-0.861) **0.881** (0.879-0.884)
							NEWS2 **0.830** (0.818-0.841) **0.843** (0.840-0.847) **0.867** (0.864-0.869)
Fernando et. al. [43]	2019	Prognostic accuracy of the Hamilton Early Warning Score and the National Early Warning Score 2 among hospitalised patients assessed by rapid response team	5491	Retrospective database analysis	Admissions to 2 hospitals in the Ottawa Hospital Network	Outcomes: In hospital Mortality And ICU admission from RRT patients	AUROC (95%CI) AUC for mortality (95%CI) HEWS 0.760 (0.75-0.7 NEWS 0.723 (0.71-0.7 Prognostic accuracy for ICU admission: **HEWS ≥3** **HEWS ≥5** **NEW** **Specificity** **63.3** (61.6-65.1) **68.4** (66.6-70.0) **64.5** (62.7 **Sensitivity** **91.8** (90.2-93.1) **76.4** (74.1-78.6) **83.4** (81.4 **Number needed to examine** **1.81** (1.76-1.85) **1.84** (1.79-1.89) **1.86** (1.81

HEWS= Hamilton Early Warning Score; NEWS2= National Early Warning Score version 2

Single parameter track and trigger scores

Cretikos et. al. [60]	2007	The objective medical emergency team criteria: A case-control study	450 cases 520 matched controls	Prospective case-control study with post-hoc analysis	MERIT study control hospitals- December 2002- May 2003	Outcomes: Unexpected cardiac arrest; unplanned admission to ICU; Unexpected death	MERIT score and MERIT score with additional criteria for threatened airway, seizures, low heart rate, low respiratory rate:				
								Sens.	**Spec.**	**AUC**	**(95%**
							Composite outcome				
							MERIT	49.1	93.7	0.71	-
							Modified MERIT	50.4	93.3	0.72	-
							Unplanned ICU				
							MERIT	56.6	93.6	0.75	(0.72 0.78)
							Modified MERIT	58.6	93.6	0.76	(0.73 0.79)
							Unexpected cardiac arrest				
							MERIT	34.8	94.0	0.64	(0.58 0.70)
							Modified MERIT	34.8	92.7	0.64	(0.58 0.69)

Smith et. al [61]	2008	A review, and performance evaluation, of single-parameter "track and trigger" systems	9987 patients- 1x admission vital sign set from each	Retrospective application of 31 scores to previously collected vital signs	58 bedded medical admissions unit May- December 2006	Vital signs at admission linked to outcome 1° outcome: in hospital mortality	**Score**	**Sens**	**Spec**	**PPV%**	**NPV%**
							Bell 2004	15.1	95.2	22.4	92.5
							Ball 2002	28.3	88.0	17.7	93.1
							Parissopoulos 2005	32.6	88.3	20.3	93.5
							Hickey 1998	24.7	91.0	21.8	93.0

Both single Parameter and aggregate weighted scores studied

Churpek et. al. [62]	2013	Risk Stratification of Hospitalized Patients on the Wards	59,643 admissions	Retrospective cohort study Aggregate weighted: MEWS; SEWS ViEWS; eCART Single parameter: MERIT Modified MERIT	Single urban academic centre in the United States	Vital signs linked to outcome Outcomes: Cardiac arrest; ICU transfer; Mortality; Composite	Area under receiver operating curve values:

Score	Cardiac arrest (95%CI)	ICU (95%CI)	Mortality (95% CI)	C o (9 C
MEWS	**0.76** (0.71-0.81)	**0.74** (0.73-0.75)	**0.87** (0.84-0.89)	**0** ((0
SEWS	**0.76** (0.71-0.81)	**0.75** (0.74-0.76)	**0.88** (0.86-0.90)	**0** ((0
ViEWS	**0.77** (0.72-0.82)	**0.73** (0.72-0.75)	**0.88** (0.86-0.90)	**0** ((0
eCART score	**0.83** (0.79-0.86)	**0.77** (0.76-0.78)	**0.88** (0.86-0.90)	**0** ((0
MERIT	**0.63** (0.59-0.68)	**0.64** (0.63-0.65)	**0.74** (0.71-0.76)	**0** ((0
Modified MERIT	**0.69** (0.65-0.74)	**0.69** (0.68-0.70)	**0.79** (0.76-0.81)	**0** ((0

Green et. al [63].	2018	Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients	107 868 patients	Retrospective database study Aggregate Weighted-NEWS; eCART Single parameter-BTF	Admissions to 5 hospitals in Illinois 2008-2013	Outcomes: In hospital death Transfer to ICU Cardiac arrest within 24 hours of observation set	**Score**	**Death in hospital AUC (95%CI)**	**IHCA AUC (95%CI)**	**ICU transfer AUC (95%CI)**
							BTF	**0.716** (0.713-0.719)	**0.602** (0.597-0.608)	**0.592** (0.591-0.593)
							NEWS	**0.777** (0.775-0.780)	**0.695** (0.689-0.700)	**0.647** (0.646-0.648)
							ECART	**0.840** (0.838-0.842)	**0.806** (0.802-0.810)	**0.723** (0.722-0.724)

Abbreviations: BTF, Between the flags; NEWS, National Early Warning Score 1; ECART, Electronic Cardiac Arrest Risk Triage score, SEWS, Standardised Early Warning Score; MEWS, Modified Early Warning Score- Subbe et. al. 2001; MADDS, Mater Adult Deterioration System; ADDS; Adult Deterioration Detection System.

1.3.3 Recognising stability

The statistical evaluation described in Table 1-4 can be used to refine scores with the aim of optimising design. However, prospective studies are necessary to determine actual, rather than projected impact of an intervention. In clinical practice, the retrospective application of a score does not account for patients who met escalation criteria but did not require a change in management, or patients in whom timely intervention would not change trajectory. In order to address the first of these issues, one study prospectively followed patients vital sign trajectories throughout admission [2]. It revealed that 66.7% of abnormal observations resolved, or the patient went on to be discharged without any intervention. A second study looked at the fate of patients meeting thresholds for escalation using the Irish National Early Warning Score [4]. In 481 medical admissions, 87 reached a NEWS of ≥ 7. Of these, 51 (64.6%) had no change in clinical management for their first episode, of whom 48 (94.1%) were discharged home, the remaining three were transferred to another hospital or died expected deaths.

This highlights a key issue of EWS summarised by one of the authors of both the VitalpacTM early warning score and the National Early Warning Score, 'neither ViEWS nor NEWS was designed to predict a change in clinical management, merely to identify a patient was sick' [64]. There are other considerations when using NEWS. Firstly, NEWS and its successor NEWS2 do not differentiate chronic from acute abnormalities, and clinical discernment is required to ascertain whether a patient in fact needs either medical review or intervention. This is particularly the case in the setting of chronic underlying changes caused by chronic disease or polypharmacy which have the potential to alter both the baseline observations and response to further pathological insult or treatment.

Secondly, an intervention at the point of triggering may not change outcome. The only prospective stepped wedge trial to introduce NEWS in place of a standard observation chart, alongside a rapid response system, to multiple hospitals failed to show significant mortality benefit (Hospital mortality 13.7 per 1000 admissions in the control phase versus 14.1 following intervention, $p=0.170$). Patients in the intervention phase had more observations documented in the 24 hours leading up to serious adverse event ($p<0.001$), however overall compliance with NEWS protocol was only 47.7% [65, 66]. Future stepped wedge studies of this type are unlikely to be possible as the majority of hospitals in developed countries now employ an EWS.

1.3.3.1 Additional inputs to aid discrimination in early warning systems- Between the Flags

If not all patients who trigger a response based on physiological criteria go on to deteriorate, are there other parameters that can be used to predict deterioration? The studies in table 1-4 analyse the sensitivity, specificity, negative predictive value, positive predictive value and area under the receiver operating characteristic curve of scores in predicting serious adverse event based on retrospective analysis of linked vital sign data. They predict what would happen if every patient had no underlying changes to their physiology and followed a specific pathway perfectly, and in addition that every member of staff reacted exactly as proscribed, without further cues or inputs into the system. By this measure, table 1-4 demonstrates that single parameter track and trigger scores are less able to statistically separate patients who are going to deteriorate from those who are going to remain stable or improve than the best performing aggregate weighted scores. However, in Australia and New Zealand where these scores are most commonly deployed, they form only part of the system that is used to escalated care.

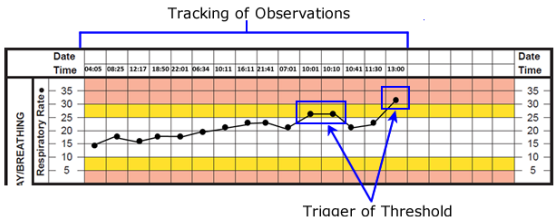
The original criteria for triggering a MET to attend a deteriorating patient, subsequently replaced by Between the Flags (see figure 1-3), include metabolic measures such as pH and lactate, unexpected drain output, uncontrolled pain and family or nurse concern. Studies analysing reasons for MET activation have reported a particular emphasis on the role on nurse concern. In the hospital that introduced the first MET, a review of activations after 8 years of use found 39% of calls involved some element of nurse concern, with 12% having no marked physiological derangement evident at the point of referral [67]. The multicentre MERIT study mirrored these findings, with 39% of calls in the hospitals introducing a MET triggered by concern [68]. Therefore pure statistical analysis of the physiological criteria does not account for patients who would be escalated by these systems despite not meeting threshold levels of vital sign derangement.

Figure 1- 3 Between the flags calling criteria taken from the Standard Adult General Observation Chart [68]

Between the Flags Calling Criteria:

Response according to

Standard vital signs calling criteria
+/-
Yellow or Red zone criteria



Yellow Zone Response

IF YOUR PATIENT HAS ANY YELLOW ZONE OBSERVATIONS OR ADDITIONAL CRITERIA* YOU MUST

1. Initiate appropriate clinical care
2. Repeat and increase the frequency of observations, as indicated by your patient's condition
3. Consult promptly with the **NURSE IN CHARGE** to decide whether a **CLINICAL REVIEW** (or other CERS) call should be made

Consider the following:

- What is usual for your patient and are there documented 'ALTERATIONS TO CALLING CRITERIA'?
- Does the trend in observations suggest deterioration?
- Is there more than one Yellow Zone observation or additional criterion?
- Are you concerned about your patient?

IF A CLINICAL REVIEW IS CALLED:

1. Reassess your patient and escalate according to your local CERS if the call is not attended within 30 minutes or you are becoming more concerned
2. Document an A-G assessment, reason for escalation, treatment and outcome in your patient's health care record
3. Inform the Attending Medical Officer that a call was made as soon as it is practicable

***Additional YELLOW ZONE Criteria**

- Increasing oxygen requirement
- Poor peripheral circulation
- Excess or increasing blood loss
- Decrease in Level of Consciousness or new onset of confusion
- Low urine output persistent for 4 hours (< 100mLs over 4 hours or < 0.5mL/kg/hr via an IDC)
- Polyuria, in the absence of diuretics (urine output > 200mL/hr for 2 hours)
- Greater than expected fluid loss from a drain
- New, increasing or uncontrolled pain (including chest pain)
- Blood Glucose Level < 4mmol/L or > 20mmol/L with no decrease in Level of Consciousness
- Ketonaemia > 1.5mmol/L or Ketonuria 2 + or more
- **Concern by patient or family member**
- **Concern by you or any staff member**

Red Zone Response

IF YOUR PATIENT HAS ANY RED ZONE OBSERVATIONS OR ADDITIONAL CRITERIA# YOU MUST CALL FOR A RAPID RESPONSE (as per local CERS) AND

1. Initiate appropriate clinical care
2. Inform the **NURSE IN CHARGE** that you have called for a **RAPID RESPONSE**
3. Repeat and increase the frequency of observations, as indicated by your patient's condition
4. Document an A-G assessment, reason for escalation, treatment and outcome in your patient's health care record
5. Inform the Attending Medical Officer that a call was made as soon as it is practicable

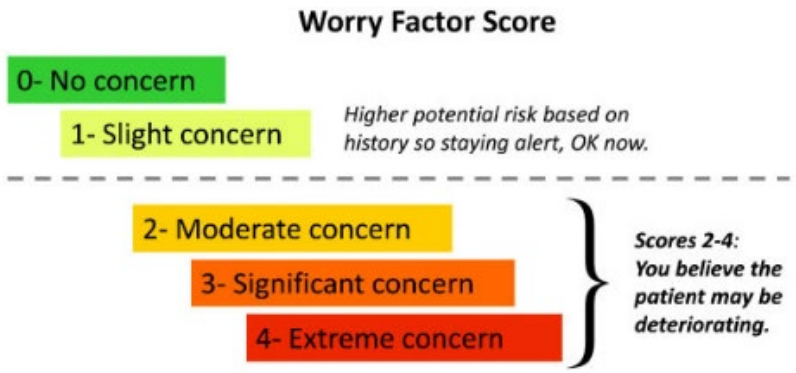
#Additional RED ZONE Criteria

- **Cardiac or respiratory arrest**
- **Airway obstruction or stridor**
- **Patient unresponsive**
- Deterioration not reversed within 1 hour of Clinical Review
- Increasing oxygen requirements to maintain oxygen saturation > 90%
- Arterial Blood Gas: PaO_2 < 60 or $PaCO_2$ > 60 or pH < 7.2 or BE < -5
- Venous Blood Gas: $PvCO_2$ > 65 or pH < 7.2
- Only responds to Pain (P) on the AVPU scale
- Sudden decrease in Level of Consciousness (a drop of 2 or more points on the GCS)
- Seizures
- Low urine output persistent for 8 hours (< 200mLs over 8 hours or < 0.5mL/kg/hr via an IDC)
- Blood Glucose Level < 4mmol/L or > 20mmol/L with a decreased Level of Consciousness
- Lactate ≥ 4mmol/L
- **Serious concern by any patient or family member**
- **Serious concern by you or any staff member**

1.3.3.2 Additional inputs to aid discrimination in early warning systems- Nurse Concern

A recent prospective study of nurse concern examined the discrimination of a 5 level nurse worry indicator scale (see figure 1-4) in medical and surgical patients admitted to an academic hospital in the USA. This reported an area under the ROC curve for transfer to ICU within 24 hours of 0.964 [69], making it more accurate in predicting which patients are going to need transfer to ICU in the next 24 hours than any current early warning scores.

Figure 1- 4 Worry factor scale [69]



Nurse concern is dependent on several factors, not merely the acuity of a patient. These can include experience level of the nurse, whether they are working in their specialty area or have been moved somewhere unfamiliar for staffing reasons, their perception of staffing and experience levels and therefore baseline comfort. Several studies have tried to elucidate in more detail what constitutes patient-specific nurse worry in a more quantifiable way that can be used to effectively package their concerns [70]. This has largely been done through qualitative methods, attempting to describe cues that nurses use and exploring the process of recognition. A systematic review of the available data revealed 10 key indicators that emerged as themes throughout the literature, a summary of which is shown in table 1-5 [71]. The majority of these align with, but are not captured by, vital signs observations collected as part of routine physiological surveillance.

Table 1-5 Nurse worry indicators adapted from Douw et al.[45]

Change in breathing	Abnormal or laboured breathing; increasing respiratory rate; dyspnoea; continued or increasing use of oxygen
Change in circulation	Cool peripheries, change in colour from patient's usual, arrhythmia
Temperature	Rigors, clammy, flushed
Change in mentation	More vague, slower than usual. Withdrawn or lethargic
Agitation	Unsettled, restless, anxious, frequently wanting to change position
Pain	New, increasing, uncontrolled or unexplained pain, particularly in chest or head.
Unexpected trajectory	Not progressing along expected recovery pattern based on diagnosis and current treatment, not eating, abdominal distension, bleeding
Patient indicates feeling unwell	Feeling of impending doom, not feeling right, unable to explain, scared when not normally.
Subjective nurse observation	Patient looks unwell. Cannot settle, change in mood, quieter, reduced motivation
Knowing without rationale	Gut feeling, intuition, something just doesn't look right

The authors of this systematic review used their findings as a basis for creating the Dutch-early-nurse-worry-indicator-score (DENWIS). The nine DENWIS indicators were then deployed in a surgical environment alongside a binary worried versus not worried option. Area under the ROC curve for composite outcome of ICU admission or mortality based on presence of DENWIS indicators was 0.85. Crucially nurse worry appeared to present earlier in the clinical trajectory than threshold derangement of EWS. In patients with an EWS of 0 but went on to have an event in 24 hours, 75% had positive DENWIS indicators, versus 13.6% of those who didn't go on to deteriorate. This was also reflected in those with EWS 1-3 where 81% of those who went on to deteriorate had positive DENWIS indicators versus 27% who remained stable or improved. This indicates the potential for escalating patients based on nurse concern who might otherwise have slipped under the radar and gone on to deteriorate.

1.4 Afferent Limb- Early Warning Scores in the setting of chronic disease- Respiratory as a paradigm

The fundamentally different physiology seen in pregnant women and children that makes NEWS a poor screening tool in these groups is also a factor in patients who have adapted to living with an underlying chronic disease. Because chronic disease frequently causes pathological alteration in baseline physiology and response to acute illness, there is an argument that these patients should be monitored differently. This is particularly the case in with the setting of chronic respiratory disease, where the adaptations caused by impaired gas exchange makes titration of oxygen to the targets set out in NEWS potentially dangerous. It is an issue acknowledged by the Royal College of Physicians. In the report of the National Early Warning Score Development and Implementation Group (NEWSDIG) that accompanied the release of the first National Early Warning Score it was stated that 'the chronically disturbed physiology of some patients with chronic obstructive pulmonary disease (COPD) could influence the sensitivity of the NEWS, which should be recognised when interpreting NEWS in these patients' [72]. A separate profile was not included in the first iteration of NEWS as NEWSDIG felt it safer to preserve the sensitivity of the score and instead advised that respiratory patients should be reviewed on a case by case basis.

In order to manage this problem, many hospitals who otherwise implemented NEWS created the addition of a profile for patients with, or at risk of, hypercapnic respiratory failure with a target saturation range of 88-92%. With similar weighting for saturations outside of this range. When these alterations are applied to a vital signs database there is a statistically lower sensitivity and positive predictive value [54, 73]. However, this statistical mismatch did not appear to lead to failure to escalate patients who went on to deteriorate and the reduction in alerts would potentially lessen the workload for clinical staff and reduce the possible effects of alert fatigue [74], while the targeted saturation range reduces the potential for harm from over-oxygenation. Although this remains a controversial point to some, the RCP responded to widespread calls for change and NEWS2 included a second profile for patients with hypercapnic respiratory failure, scale 2, setting target saturations in this group for patients requiring oxygen at 88-92% (See figure 1-5) [75].

Figure 1- 5 NEWS2 including scale 2 for patients with hypercapnic respiratory failure [76]

Physiological parameter	Score 3	2	1	0	1	2	3
Respiration rate (per minute)	≤8		9–11	12–20		21–24	≥25
SpO_2 Scale 1 (%)	≤91	92–93	94–95	≥96			
SpO_2 Scale 2 (%)	≤83	84–85	86–87	88–92 ≥93 on air	93–94 on oxygen	95–96 on oxygen	≥97 on oxygen
Air or oxygen?		Oxygen		Air			
Systolic blood pressure (mmHg)	≤90	91–100	101–110	111–219			≥220
Pulse (per minute)	≤40		41–50	51–90	91–110	111–130	≥131
Consciousness				Alert			CVPU
Temperature (°C)	≤35.0		35.1–36.0	36.1–38.0	38.1–39.0	≥39.1	

Following the introduction of NEWS2 there has been discussion regarding the scope of scale 2, as its use is currently only mandated in those patients with COPD who also have evidence of high CO2 on blood gas. A widely cited cluster randomised control trial whereby all patients being transferred to hospital by ambulance with either a diagnosis of COPD, or appropriate smoking history, were allocated to usual high flow oxygen therapy or controlled oxygen therapy targeting saturations of 88-92%, reported in-hospital mortality of 9% in control versus 4% in intervention. The mean pre-hospital treatment time was 47 minutes and the intervention group had a relative risk of in hospital mortality 0.42 (95% confidence interval 0.20 to 0.89; P=0.02). [76]

NEWS2 has also been retrospectively applied to the DECAF database, a cohort of 2645 consecutive admissions of patients with exacerbation of COPD (confirmed pre-admission on spirometry) to six hospitals in the UK. Applying scale 2 to all patients instead of purely to those with confirmed hypercapnic respiratory failure, led to an absolute reduction in alert frequency of 12.6%. Although statistically this made the score less sensitive, no patient with a NEWS2 score that was downgraded to being low risk using this approach died on the same day as a downgraded observation set. [77]. It is also worth noting the Thoracic Society of Australia and New Zealand guidelines relating to acute oxygen use in

adults promote target saturations of 88-92% in patients at risk of type 2 respiratory failure, not limiting diagnosis to COPD [78].

This reduction in saturation targets is in line with growing recognition of the potential harm of hyperoxia, with a recent systematic review and meta-analysis reporting liberal oxygen delivery strategies to be associated with a significant increase in both in-hospital and 30 day mortality in comparison to conservative administration across a general hospital population [79]. While further prospective work is required, particularly relating to specific conditions such as wound healing in the post-surgical patient, this approach has the potential to reduce both harm to patients, and workload for staff. Releasing them to carry out other tasks and respond to acute clinical need.

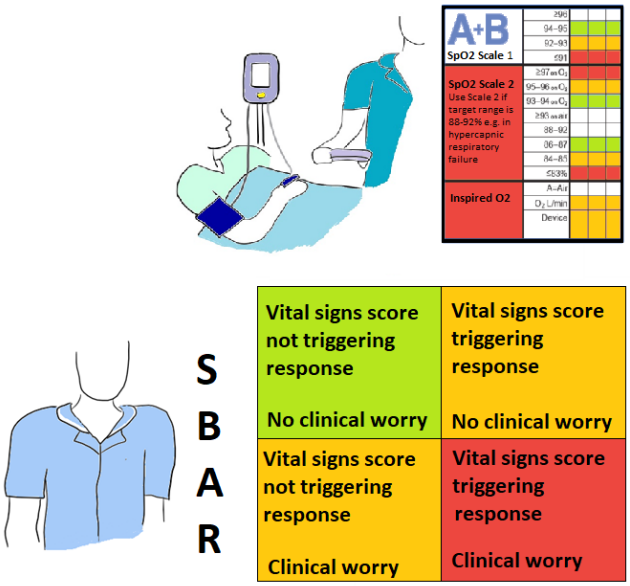
1.5 Efferent Limb- Referral-

The output of the afferent limb is the process of escalating the patient for some form of clinical response as the first stage of the efferent limb. Whether this be to generate a review from a single clinician, or from a specialist multidisciplinary team.

Despite interventions and education programmes, referral of patients often does not occur in line with early warning score protocols. Of 372 incidents reported to the National Reporting and Learning System between 1st January and 31st December 2015, 36 were not escalated despite triggering local EWS protocols (31 on ward and five in ED), and in 11 there was failure to instigate proper treatment (nine on ward and two in ED) [80]. Qualitative studies suggest that experienced nurses will rely on experience and judgement where the local protocol recommends escalation if clinically they feel the patient is stable or there is a plan in place that allows them to administer treatment [28]. As with the factors discussed surrounding the process of recognition, this is particularly the case in patients with recognised stable derangement at baseline, such as patients with respiratory disease. In one study examining out of hours escalations it was noted that the respiratory ward had the highest rate of triggers, but escalated just 12% of them. It is of note that only one of the 25 deaths recorded in the study happened within 24 hours of a trigger, and the median time from trigger score of $\geq$6 to death was 11 days [35]. Equally, as detailed in the previous section, nursing staff will often escalate patients who do not meet local protocol thresholds if their intuition suggests there is a problem, as one nurse put it, ‘Something is wrong, the obs are fine, but something is definitely

wrong' [81]. Figure 1-6 displays the possible conflict that can arise when considering both vital signs and end of the bed assessment.

Figure 1- 6 The interplay of nurse concern and early warning score values



In addition to evaluation of the clinical status of the patient, nursing staff appear to take other factors into account when it comes to physically escalating a patient. On consideration is the extent to which they feel doctors are supporting them in care of the patient [82-84]. This appears to be backed up by a higher incidence of MET calls among outliers [85], although the small number of studies examining this mean the ability to draw conclusions is limited. Factors which make it less likely that a clinician outside of the ward team or specialist rapid response team will be called include presence of more senior members of the home team, such as registrar or consultant, or a plan for management. Where there is confidence in response to the ward team they are preferentially called as it is felt they would be more familiar with the patient's condition and represent a lesser lag in commencing treatment. A fear of criticism from the home team also made nursing staff less likely to call a MET and made them confer with colleagues before making a decision, this was particularly the case for more junior nurses.

With the advent of fully integrated electronic observation and task allocation systems there is the potential for automatic escalation of threshold scores to the appropriate

clinical staff. Although there is some evidence that automatic escalation improves outcomes, possibly through reducing time to escalation, this has not been demonstrated in a ward with standard monitoring and escalation systems and therefore these findings cannot be generally assumed to be true [86]. Notwithstanding automatic escalation protocols, the introduction of electronic task allocation systems has reduced the time taken for a referral to happen. However, implementation at a university teaching hospital with tertiary services did not change adherence to protocols in terms of which patients were referred.

1.6 Efferent Limb- Response-

The second stage of the efferent limb is the clinical response to a referral. This has previously been described in terms of following either a ramp up or ramp down approach [87]. The ramp up system involves a single clinician or nurse responding to concern and assessing the patient before deciding what further resources are needed. The ramp down calls a multidisciplinary team to the bedside to review the patient, with those elements that are not required then leaving as part of a de-escalation.

Although in the past these two methods have been distinct from one another, in practice, most systems now have elements of both approaches dependent on the perceived level of acuity of the patient being referred as part of a two tier system of response. A single clinician is deployed as a first stage where clinical review is indicated but patient acuity is not felt to immediately warrant an emergency or cardiac arrest team call. For example, in NEWS2 this occurs at a score of 5-6, or as a 'Yellow Zone Response' in the Australian Between The Flags system. The intention in triggering clinical review at an earlier stage is that intervention at a lower level of acuity may prevent deterioration in a proportion of patients. As most systems do not record the outcome of these escalations it is unknown what proportion of patients who reach these trigger thresholds are reviewed, what interventions are instigated, and whether these interventions alter patient trajectory. It is therefore unclear what factors clinicians use to discriminate who needs review beyond having threshold scores and taking into account those factors discussed in relation to referral. It is likely that alert fatigue created by persistently high scores may make a patient less likely to be reviewed. Also that where workload and escalation burden do not allow for review of all patients reaching threshold scores to be reviewed, there may be patients who the clinical team are unable to address at this stage. However more work is needed to more fully understand the factors involved.

When patient acuity is deemed to be at a higher level, above 7 on the NEWS and NEWS2 protocol, it is advised that the patient should be reviewed by someone with critical care competencies. There have been two main approaches to meeting this requirement. The first is the critical care outreach team (CCOT) consisting of senior specialist nurses who can provide review, support and liaison with ICU. Delay in CCOT referral has been shown to increase both ICU admission and mortality in both medical and surgical patients [88]. Introduction of a CCOT was also shown to reduce mortality of ICU patients following discharge to the ward environment [89] This suggests their input adds value over ward management alone, in line with recommendations from a government review of critical care services promoting integration of services to create a 'critical care without walls'[1]. However presence of CCOT in those hospitals where it exists is inconsistent and not always available 24 hours a day, 7 days a week. As with any intervention, impact is also linked to staff engagement. A stepped wedge analysis of introduction to hospitals in Iran failed to show a mortality benefit where it was noted staff engagement was low [90]

At an equivalent NEWS2 score in many NHS hospitals, and in the two tier Between the Flags system in Australia as a 'Red Zone Response', a Medical Emergency Team call is triggered. In the NHS this is often in addition to the presence of CCOT who may already be involved with a deteriorating patient due to concerns raised by nursing staff. Medical emergency or rapid response teams are comprised of several different key members of a multidisciplinary nature. The common components include a medical registrar, an intensive care representative and more junior members of medical staff to assist with investigation and management. Additional members during office daytime hours often include a bed manager, pharmacist and resuscitation officer.

The largest study examining implementation of MET was the multicentre MERIT study (see table 2). Primary analysis did not suggest significant improvement in mortality between the intervention and control sites, despite significantly increased emergency team activation [91]. However, authors performing a post hoc analysis suggested an element of leak to the control limb, with cardiac arrest teams in these hospitals employed more like an emergency response team for the duration of the study [68], and a reduction in cardiac arrests and unexpected deaths seen between baseline and intervention at control and intervention sites [91]. Other studies have suggested reduction in cardiac arrest rates, however these were largely before and after without a contemporaneous control group, with the inherent drawbacks of this study design.

Having introduced a National Score in order to standardise the afferent pathway, there is a similar argument for introducing a standardised efferent pathway, i.e. familiarity and ability to assess and share good practice. The most integrated system with a standardised afferent and efferent pathway is seen in New South Wales. Significant clinician feedback was used to develop an afferent pathway with multiple sources of input, Between the Flags, and linked MET. This was then introduced across all 225 public hospitals in NSW. The introduction of a two tier system using the yellow and red indicators previously described in place of an all or nothing MET system has seen a decrease in ICU mortality following medical review suggesting that patients are being identified earlier and in doing so a change in trajectory is possible [92]. In addition, there has been a significant decrease in cardiac arrest of 42% compared to baseline pre-introduction ($p<0.05$) [93]. Data describing not for resuscitation decisions are not available. However, even if a proportion of the improvement in cardiac arrest rates is due to appropriate escalation plans being made in a higher proportion of patients, this remains an important patient outcome.

It is unclear which category of response provides the greatest benefit, or whether they should be used concurrently, as studies have examined the impact of introducing CCOT or a MET where previously there was only an arrest team, but no work exists to provide a direct comparison or follow outcomes through transition between these systems. The fundamental consequence of both approaches is to deliver greater resources and experience to a potentially deteriorating patient with the aim of changing trajectory or moving to a higher level of care at a lower level of clinical complexity.

1.7 Summary

This literature review has explored the evidence base relating to the challenges observed at each stage of the process of recognising and responding to a deteriorating patient. It has highlighted significant current gaps in the literature. These include an understanding of the characteristics of hospital inpatients with specific focus on understanding trajectory, and to what extent it is possible to alter this; how NEWS2 impacts on both recognition of the deteriorating patient and the staff using it; and finally how changes in physiology created by chronic disease states impact the predictive and clinical value of the score.

There is also a need to understand where deviation from protocol is due to human error or alert fatigue, and where other information not currently understood or captured is being integrated to assess the patient. Vital signs observations are not a stand-alone measure. Cues provided by the patient they are taken from are taken into account in other healthcare systems and with further evidence to support this could be successfully deployed in an overt, standardised manner in the NHS in order to further improve communication at the clinical interface.

1.8 Research Questions Generated

This PhD plans to contribute further through answering the overarching research question of, 'How effective are early warning scores in highlighting deterioration across the hospital?' This will be done through addressing the following aims :

- Examine escalations as a surrogate measure of workload generated by NEWS2 to determine whether there is a clear relationship between workload and patient deterioration:
    - Looking at patterns by specialty
    - Examining in versus out of hours escalations
    - Determining workload created in comparison to patient deterioration- i.e. number needed to escalate before an outcome is seen.
- Profile patients who go on to experience in hospital death versus those who don't to determine whether it is possible to differentiate between these groups:
    - First score, average score, final score;
    - Further information including morbidities; frailty; age- all routinely documented;
- Perform regression analysis to determine whether change in NEWS2 can be used more effectively to predict deterioration, or equally importantly, stability.
- Evaluation of what is used in addition to vital signs to trigger a response- particular focus on the role of nurse concern in either escalating patients without deranged vital signs, or not escalating patients with deranged vital signs through analysis of critical decision methods interviews.

  Postulation of a set of nurse concern criteria to be developed for use alongside NEWS2 with the aim of recognising deterioration at an earlier point in the patient journey and improving recognition of stability.

2 Data for Quantitative Studies

As part of this thesis three retrospective data studies were carried out. These all used the same outcomes-linked data-set with the details provided below. All analysis was carried out using STATA 17.

2.1 Setting for all studies

The setting, Nottingham University Hospitals Trust (NUHT), consists of two hospitals with between 1500-1700 overnight beds, depending on demand. NUHT is a regional referral centre for Neurosurgery, Cardiac and Thoracic Surgery as well as several medical specialties. In addition, it has a level one trauma centre ensuring a consistent flow of admissions, requiring flexible capacity, and placing significant demands on staff.

2.2 Data Source

In the summer of 2014, an electronic observations software system (Nervecentre) enabling vital signs observations to be inputted at the bedside using handheld devices was deployed across the hospital. The same software collated the vital sign data into an early warning score. This local early warning score (LEWS) was similar to the later NEWS2 score (see Table 2.1 for comparison) but had scoring cut points that allowed slightly more deranged vital signs before scoring thresholds were reached, a graded score for both inspired oxygen and AVPU, the inclusion of urine output, and the exclusion of oxygen saturations. Another deviation from NEWS2 was the presence of different profiles to adjust for changes to baseline physiology in the setting of chronic disease. For example, being anuric for a period of more than 18 hours would score 3 on LEWS, but a chronic anuria profile was available for patients on renal replacement therapy to avoid unnecessary escalation. If the LEWS was elevated beyond set threshold scores the software system automatically prompted clinical intervention/escalation through requests for medical review, transmitted to the clinical staff via a mobile phone. In June 2019 the software was amended to employ the NEWS2 algorithm; automatic escalation at set thresholds continued as dictated by the NEWS2 protocol published by the Royal College of Physicians [1]. The combined system captures all vital sign data, early warning scores, and automated requests for clinical intervention/escalation.

This software system was used to create a database of vital signs observations, including heart rate, blood pressure, respiratory rate, oxygen saturation, fraction of inspired

oxygen, temperature, conscious level and urine output, linked to outcome and demographics, admitted between 1st January 2016 and 31st December 2019, following Health Research Authority and Information Governance approvals.

The NEWS2 score was calculated retrospectively for each set of vital signs observations collected prior to the introduction of NEWS2 in 2019. Cut points were applied in line with the escalation protocol published with NEWS2 in which a score of 5 or more dictates an urgent response and hourly monitoring and 7 or more an emergency response with continuous monitoring. [94]

Table 2-1 Scoring parameter for NEWS2 versus Local Early Warning Score in place prior to June 2019

	3	2	1	0	1	2	3
Resp Rate NEWS2	≤ 8		9-11	12-20		21-24	≥25
Resp Rate LEWS	≤ 8		9-11	12-20		21-29	≥30
Sats NEWS2 Scale 1	≤91	92-93	94-95	≥96			
Sats NEWS2 scale 2	≤83	84-85	86-87	88-92 & ≥93 on air	93-94 on oxygen	95-96 on oxygen	≥97 on oxygen
Sats LEWS	Oxygen Saturations not included in LEWS						
FiO2 NEWS2		On Oxygen		Air			
FiO2 LEWS				Air	40-60%		≥61%
BP NEWS2	≤90	91-100	101-110	111-219			≥220
BP LEWS	≤80	81-90	91-100	101-220			≥221
HR NEWS2	≤40		41-50	51-90	91-110	111-130	≥131
HR LEWS	≤40	41-50		51-100	101-110	111-129	≥130
AVPU NEWS2				Alert			CVPU
AVPU LEWS				Alert	V	CP	U, New neurology
Temperature NEWS2	≤35.0		35.1-36.0	36.1-38.0	38.1-39.0	≥39.1	
Temperature LEWS			35.1-36.0	36.1-38.0		38.1-39.0	≥39.1
Urine Output NEWS2	Urine output not included in NEWS2						
Urine Output LEWS (ml/hr in catheterised patients)	Not Passed in 18 hours/ Anuric not on RRT or <10ml/hr	Not Passed in 12 hours or 10-20ml/hr	Not passed in 6 hr or 21-30ml/hr Or anuric on RRT	Passed in 6 hr or 31-199 ml/hr	≥200ml/hr		

*RRT=Renal Replacement Therapy

2.3 Approvals

This study was carried out as part of the Health Research Authority approval for the questions put forward by this thesis as IRAS protocol ID 270837. As the database contains anonymised data recorded in the course of the admission for the purposes of patient care and no interventions were requested it did not require Research Ethics Committee approval.

Approval from the Nottingham University Hospitals Caldicott guardian, Information Governance Department and Research and Innovation department were obtained with all regulatory paperwork included in the Appendix.

2.4 Data Definitions

Term	Definition
Observation	Set of vital signs recorded at bedside. Taken together each set amalgamated to NEWS2 score
Vital Sign Score	Indicates how far each vital sign deviates from set normal range, calculated at collection of each observation with a weighting of 0-3.
NEWS2 score-	National Early Warning Score version 2- Published by the Royal College of Physicians (RCP) in 2017 and mandated for use across the NHS in the UK. NEWS2 is an aggregate early warning score. NEWS2 is a continuous variable from a minimum score of 0 to a maximum of 20.
Scale 1	Oxygen target scale for NEWS2- for use in patients with no evidence of type 2 respiratory failure. Target saturations 94-98%.
Scale 2	Oxygen target scale for NEWS2- for use in patients with evidence of type 2 respiratory failure- used in all patients with diagnosis of COPD in line with clinical practice. Target saturations 88-92%.
Cut points-	NEWS2 scores at which certain actions are advised as per the protocol published by the RCP: **NEWS2 of 5-6**: Minimum hourly observations, registered nurse to immediately inform medical team and request urgent assessment within 1 hour by clinician with core competencies in care of acutely ill patients. Provide clinical care in an environment with monitoring facilities. **NEWS2 of less than 5 but3 in one category:** Separate category in original scoring protocol but clinically treated the same as a score of 5 or more. **NEWS2 of 7 or more**: Continuous monitoring of vital signs; registered nurse to inform registrar or above in medical

	team, emergency assessment within 30 minutes by team with critical care competencies and advanced airway management skills; consider transfer to level 2 or 3 area with clinical care in an environment with monitoring facilities
Number Needed to Evaluate [95-97]	A measure of demand used as a surrogate for workload in this and other studies. $NNE = \frac{FP + TP}{TP} = \frac{1}{PPV}$ FP: Number of observations exceeding threshold for escalation not resulting in death within 24 hours TP: Number of observations exceeding threshold for escalation resulting in death within 24 hours PPV: Number of true positives as a fraction of the total number of observations exceeding the threshold for escalation
Primary Outcome	Death within 24 hours of an observation set

2.5 Strengths and Weaknesses

The size and completeness of our data set (all observations were input directly onto devices at bedside with a very small percentage of missing or impossible entries) strengthens confidence in our findings. Other strengths include the fact that all elements of the NEWS2 score were incorporated in the vital signs observations sets collected and that ICD10 coding made it possible for patients to be assigned to the appropriate oxygen scale. The TRIPOD checklist [98] for reporting performance of predictive scores and the STROBE statement for reporting cohort studies were applied through design, analysis and reporting. Lastly while area under the ROC curve is the most commonly used measure applied in studies relating to predictive models such as NEWS2, it is now recognised that due to the small percentage of outcome (death) within a hospital population, area under precision recall curves give added information, so both are included where appropriate [96, 97, 99, 100].

Limitations included the fact data were retrospective and from a single centre. It was not possible to retrospectively apply scale 2 to all patient groups who might be managed using scale 2 throughout the entire study period, therefore the decision to apply to patients with a diagnosis of COPD was a pragmatic approach to ensure consistency. In addition, any vital signs coded as 'End of Life Care' (meaning all interventions would be aimed at palliation of symptoms rather than prolonging life) were removed from the

analysis as a surrogate for intention to intervene to attempt to influence clinical trajectory.

The relatively small number of outcomes also represents a higher risk of type 2 error in examining the statistical discrimination of these models. While the use of multiple vital signs from an individual care episode could at first glance appear to be a limitation, this approach has been validated in the literature [101] and has become a recognised approach to evaluating early warning scores [57, 102-104].

3 Understanding demographics- The effect of implementing the NEWS2 escalation protocol in a large acute NHS trust: a retrospective cohort analysis of mortality, workload and ability of early warning score to predict death within 24 hours

3.1 Introduction

The NHS is facing unprecedented challenges with rising admissions and increasing patient need on a background of finite resources and staffing challenges all heightened by the demands due to the COVID pandemic. Understanding hospital workload is key to managing resource allocation to provide safe and efficient patient care. The literature review described in the previous chapter outlines the current research base pertaining to the development of early warning scores and defined the questions this thesis seeks to answer. The focus of the study outlined in this chapter was to provide context for later studies presented by describing the characteristics of the population being studied and the impact of introducing NEWS2 into that population

Although there is a single universal early warning score deployed in all adult inpatients outside of maternity, a hospital population is not homogenous. In order to understand the ability of NEWS2 to predict clinical deterioration, and the consequent workload generated by escalation of threshold scores as dictated by the associated protocol, the background mortality rate and variation within different populations of inpatients needs to be explored. This is vital as both the positive predictive and negative predictive values, and therefore potential workload implications of early warning scores, depend on the mortality rate (i.e. prevalence) in the population in which they are being used. Any changes to NEWS2 on its revision in 2023 are likely to have a significant impact on future hospital workload and staffing needs in the NHS. We set out to describe the inpatient population and explore any differences in characteristics of the two cohorts described; to establish whether there were any changes in admission numbers and mortality in a large teaching hospital over a 4-year period; and to assess the impact of applying 2 different

early warning scores, including analysis of the real time introduction of NEWS2 and its associated protocol, on workload.

3.2 Methods

3.2.1 Design

This study consisted of retrospective cohort analyses of medical and surgical inpatient admissions.

3.2.2 Data source

The entire database described in the previous chapter was analysed in this study. All observation sets recorded between 1st January 2015 and 31st December 2019 for adult admission episodes were included in the analysis.

3.2.3 Statistical analysis

Trends in admission rates, mortality, length of stay and early warning scores were analysed, both within the whole population and subpopulations of surgical and medical specialties defined by admission specialty. Patients admitted to specialised day-case areas were removed from the analysis as these represent a very different population subset and are managed in a different manner.

3.2.4 Defining risk factors for mortality and length of stay

Univariate logistic regression analysis was used to identify which variables (from those available at admission) had a significant association with in-hospital mortality or longer length of stay, beyond the population median. Variables assessed included age, gender, discharge from hospital in the preceding 30 days, NEWS2 score at admission and time, day and month of admission. These were built into multivariable models. Variables that remained significant for inclusion were inputted into the final models in order to illustrate differences in behaviour between the two cohorts.

3.2.5 Evaluation of workload

System demand was used as a surrogate for workload (defined as request for medical review, or escalation) as protocoled by the early warning scoring system was assessed in several ways.

3.2.5.1 Recorded escalations

Data on the number of actual requests for review/escalations were measured both pre (using LEWS) and post the change to NEWS2 in June 2019.

3.2.5.2 Observations reaching threshold score for escalation

In addition to recorded activity, predicted escalations based on the scoring thresholds for both early warning scores used during the study period were calculated. An assumption that all scores reaching the protocoled threshold for intervention/escalation would have led to an intervention was made for the purposes of this study and any observation with a score at or above that thresholds was considered a predicted positive. i.e. for LEWS all scores of ≥4 were counted as flagged to a junior doctor and scores of ≥6 counted as flagged to registrars; and for NEWS2 all scores of ≥5 were counted as being flagged to a junior doctor and scores of ≥7 as being flagged to registrars.

Where the specified score was not in use at the time of the observation set being recorded, as was the case with NEWS2 before June 2019 and LEWS score after June 2019, the score was calculated from the component vital signs. NEWS1 was also calculated to determine the level of difference when compared to NEWS2. The individual components to calculate NEWS2, including the new confusion component of ACVPU, were all recorded as part of the system prior to its introduction in June 2019 and are therefore present consistently in the dataset. When applying oxygen saturation targets as part of NEWS2, scale 2 was employed in patients with a diagnosis of COPD and scale 1 to those without. Although this is not in line with Royal College of Physicians guidance for the use of scale 2 in NEWS 2 there is precedent in the literature that applying target saturations of 88-92% in all patients with a diagnosis of COPD requiring oxygen improves outcome [3]. In

addition, this allowed for a consistent approach both before and after the introduction of NEWS2.

2.2.5.3 Analysis of score performance

These data were used to assess the performance of the score across the different cohorts. This included calculation of sensitivity and specificity at the protocol cut points applicable to each score, i.e. 4 and 6 for LEWS and 5 and 7 for NEWS2, as well as area under the ROC curve for the score as a whole for identifying the outcome of death within 24 hours of an observation.

The final metric calculated was the number needed to evaluate (NNE or workup detection ratio). NNE has been proposed as a method for comparing the ability of early warning scores to accurately predict clinical deterioration in the context of the workload they generate and provide an indirect measure of the cost-efficiency of each alert and of the early warning score employed. In this paper, the method of Kipnis et al. was used, and is defined as the number of observations that it is necessary to respond to in order to pick up one outcome of death within 24 hours [95-97]

$$NNE = \frac{FP + TP}{TP} = \frac{1}{PPV}$$

Here false positives refer to observations reaching threshold not followed by outcome of death within 24 hours and true positives refer to observations reaching threshold that were followed by death within 24 hours.

It should be noted that in both LEWS and NEWS2 the nursing and clinical staff looking after a patient have the ability to pause escalation or adjust parameters based on the clinical situation. Therefore, the number of recorded escalations is expected to be lower than the number of observations which met the threshold for escalation.

All analyses were carried out in STATA 17. Approval was given by the UK Health Research Authority (IRAS ID 270837) and Nottingham University Hospitals Trust's Caldicott guardian, Research and Innovation team and Information Governance department (Ref: DG20-000049-D and IG0025). As the study did not involve human participants and was limited to routinely collected data anonymised prior to extraction, the HRA did not require research ethics committee review.

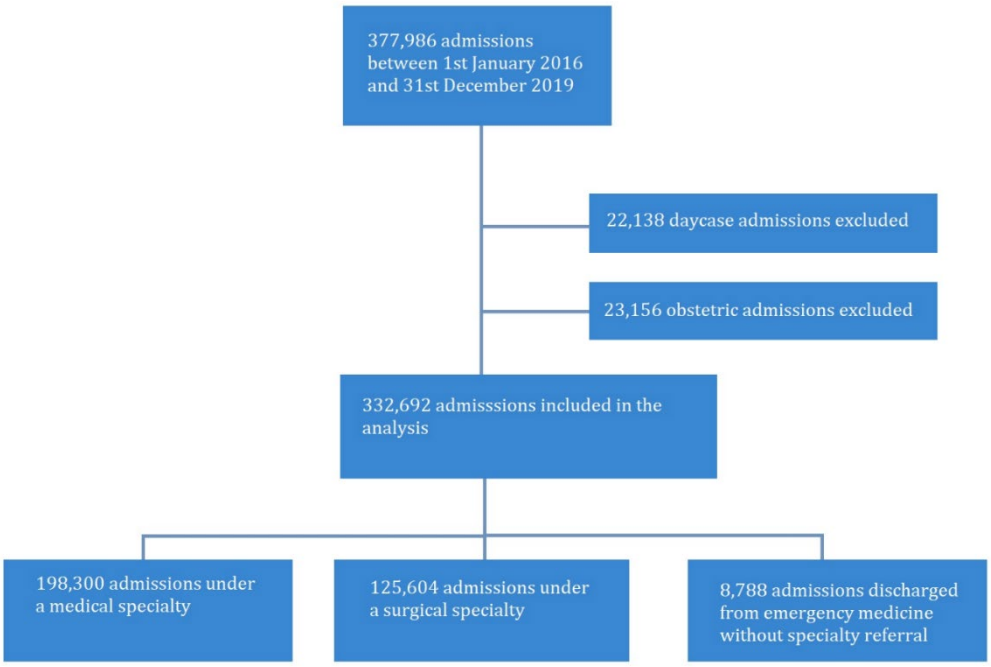
In addition to these analyses, a freedom of information request was sent to all Acute Trusts in the NHS in England to further determine the applicability of these findings. The information requested included whether the Trust employed electronic observations systems, what platforms were used, whether automatic escalations were triggered at threshold scores and what those thresholds were.

3.3 Results

3.3.1 Admissions and length of stay

332,682 adult patients were admitted between 1st January 2016 and 31st December 2019 (Figure 3-1). This excluded 23,156 patients admitted under obstetrics who were managed using a different scoring system and 22,138 patients admitted to day-case units. Median age at admission was static at 61 years throughout the four-year study period. 8,788 (2%) were discharged from the emergency department without specialty referral and were included in total numbers but not analysed separately. 198,300 (60%) were admitted under medical specialties and 125,604 (38%) under surgical specialties (figure 3-1).

Figure 3-1 Consort diagram of admissions during study



Admissions rose by 19% from 76,055 in 2016 to 90,587 in 2019. Total bed days rose 10% from 433,382 to 477,485. Readmissions also rose- 4.6% of admissions in 2019 had been discharged in the preceding 30 days compared to 3.5% in 2016, accounting for 5.3% of surgical admissions and 3.8% of medical admissions. There was a small decrease in median length of stay in patients under medical specialties from 2.1 days in 2016 to 1.9 days in 2019. Length of stay was static amongst patients admitted under surgical specialties as shown in Table 3-1.

The reduction in length of stay seen in patients under a medical specialty was partly attributable to a reduction in bed days by patients who had been declared medically fit for discharge and were waiting placement (a bed in a care home, or social service input). The equivalent of 174 beds were occupied for an entire year in 2016, and 156 in 2019 by patients who had been declared medically fit for discharge; even at the lower level seen in 2019 this equates to almost six 28 bedded wards being occupied for a whole year by patients ready for discharge. 63% (97) of these bed years were accounted for by patients aged over 75.

Table 3-1 Hospital admissions- Total, Medicine and Surgery- between 2016 and 2019

		2016	2017	2018	2019
Admissions	Total	76,055	81,379	84,671	90,587
	Surgery	29,120	31,006	31,890	33,588
	Medicine	46,182	49,034	51,483	51,601
Admissions within preceding 30 days (% of total admissions)	Total	2646 (3.5)	3274 (4.0)	3811 (4.5)	4196 (4.6)
	Surgery	1,259 (4.3)	1,537 (5.0)	1,795 (5.6)	1,780 (5.3)
	Medicine	1,353 (2.9)	1,643 (3.4)	1,951 (3.8)	1,967 (3.8)
Bed Days		453,382	451,482	456,417	477,485
Median age (IQR)		61 (46-76)	61 (46-76)	61 (41-76)	61 (41-76)
Median LOS (IQR)	Total	2.1 (0.8-6.5)	2.0 (0.7-5.9)	2.0 (0.7-6.0)	1.9 (0.5-5.7)
	Surgery	1.3(0.4-4.1)	1.3(0.4-4.2)	1.4 (0.4-4.3)	1.4 (0.4-4.2)
	Medicine	3.1 (1.0-8.2)	2.7(0.9-7.5)	2.7(0.9-7.4)	2.9 (0.9-7.5)
Mortality (% of admissions)	Total	2821 (3.7)	2853 (3.5)	2772 (3.3)	2818 (3.1)
	Surgery	335 (1.2)	352 (1.1)	319 (1.0)	347 (1.0)
	Medicine	2481 (5.3)	2493 (5.1)	2449 (4.8)	2368 (4.8)

Bed days= total beds occupied for 24 hours a day; LOS= Length of stay; Surgery= all admissions under surgical specialties; Medical = all admissions under medical specialties; Total = Surgical + Medical + Emergency department.

*Actual recorded, not predicted.

The multivariable logistic regression analysis demonstrated that across all admissions, patients were more likely to have a longer length of stay if older, female, had a NEWS2 score of five or more at admission, presented overnight or were admitted in winter (Table 3-2). In patients under a surgical specialty, having been discharged in the preceding 30 days was associated with lower likelihood of length of stay greater than two days, whereas in medicine a previous discharge within 30 days was associated with greater risk of length of stay longer than two days.

Table 3-2 Multivariate analysis for factors readily available at admission which were associated with being in hospital length of stay longer than the population median of 2 days: Analysis of total population and for division into Medical and Surgical cohorts

Total Population					
Baseline variables at admission		OR	Adj OR	95% CI	p
Age	18-40	0.31	0.33	0.32-0.33	<0.001
	41-60	0.59	0.61	0.60-0.62	
	61-75	1.00			
	76-85	1.40	1.37	1.34-1.40	
	86+	2.10	2.03	1.98-2.08	
NEWS2 >5		4.01	3.25	3.16-3.36	<0.001
Sex (Females versus Male)		0.99	1.04	1.03-1.06	<0.001
Presenting overnight (1700-0800)		1.38	1.35	1.33-1.37	<0.001
Admissions preceding 30 days		1.34	1.30	1.28-1.33	<0.001
Year of admission	2016	1.00			<0.001
	2017	0.91	0.91	0.90-0.93	
	2018	0.91	0.93	0.91-0.95	
	2019	0.89	0.91	0.89-0.93	
Admission quarter *	Dec- Feb	1.0			<0.001
	Mar- May	0.96	0.99	0.97-1.01	
	Jun- Aug	0.91	0.95	0.93-0.97	
	Sep-Nov	0.93	0.98	0.96-1.00	

Surgery					
Baseline variables		OR	Adj OR	95% CI	P
Age	18-40	0.37	0.35	0.34-0.37	<0.001
	41-60	0.66	0.63	0.61-0.65	
	61-75	1.00			
	76-85	1.27	1.26	1.21-1.30	
	86+	1.80	1.68	1.60-1.77	
NEWS2 >5		3.94	3.51	3.24-3.80	
Sex (Females versus Male)		1.24	1.29	1.26-1.32	<0.001
Presenting overnight (1700-0800)		1.57	1.64	1.60-1.68	<0.001

Admissions preceding 30 days		0.91	0.90	0.88-0.93	<0.001
Year of admission	2016	1.00			0.015
	2017	1.00	1.01	0.97-1.04	
	2018	0.98	1.00	0.97-1.03	
	2019	0.94	0.96	0.93-0.96	
Medicine					
Baseline variables		OR	Adj OR	95% CI	P
Age	18-40	0.29	0.32	0.31-0.33	<0.001
	41-60	0.56	0.59	0.58-0.61	
	61-75	1.00			
	76-85	1.43	1.44	1.41-1.49	
	86+	2.16	2.23	2.16-2.31	

3.3.2 Mortality

Mortality rates varied with season when examined monthly but remained relatively constant over the study period. However, because of the increase in overall admissions the percentage mortality decreased year on year. Mortality was significantly lower in patients discharged from a surgical specialty at 1.0-1.2%, compared to 4.8-5.3% in patients discharged from a medical specialty ($p<0.001$).

Several variables were associated with risk of mortality and were common to both medical and surgical patients (Table 3.3). Patients were more likely to die in hospital if they were older, had a NEWS2 score of five or more at admission, had been discharged in the preceding 30 days or presented in the winter. Surgical patients were more likely to die in hospital if presenting overnight or were female. Medical patients were more likely to die if male and risk of mortality in medical patients overall decreased with admission year. The degree to which each of these variables was associated with risk of mortality was different when comparing medical and surgical patients, with NEWS2 of five or more and older age associated with a higher risk of mortality in a surgical than medical population.

Table 3-3 Multivariate analysis of factors associated with a significantly higher or lower risk of mortality

Total Population					
Baseline variables		OR	Adj OR	95% CI	p-value
Age	18-40	0.07	0.09	0.08-0.11	<0.001
	41-60	0.37	0.42	0.39-0.46	
	61-75	1.00			
	76-85	1.76	1.70	1.62-1.80	
	86+	2.97	2.87	2.72-3.03	
NEWS2 >5		8.57	6.41	6.14-6.69	<0.001
Admissions preceding 30 days		1.77	1.62	1.55-1.70	<0.001
Admission quarter	Dec- Feb	1.00			<0.001
	Mar- May	0.80	0.86	0.81-0.91	
	Jun- Aug	0.77	0.87	0.83-0.93	
	Sep-Nov	0.80	0.92	0.87-0.97	
Sex (Females versus Male)		0.92	0.91	0.87-0.95	<0.001
Presenting overnight (1700-0800)		1.28	1.14	1.10-1.19	<0.001
Year of admission	2016	1.00	1.00		0.003
	2017	0.94	0.97	0.92-1.03	
	2018	0.88	0.92	0.87-0.98	
	2019	0.83	0.91	0.86-0.96	
Surgery					
Baseline variables		OR	Adj OR	95% CI	p
Age	18-40	0.07	0.07	0.05-0.12	<0.001
	41-60	0.30	0.32	0.26-0.39	
	61-75	1.00			
	76-85	2.34	2.29	1.99-2.63	
	86+	5.48	5.19	4.75-6.40	
NEWS2≥5		11.63	9.49	8.24-10.92	<0.001
Sex (Females versus Male)		1.13	1.14	1.02-1.28	0.018
Admissions preceding 30 days		1.42	1.43	1.25-1.64	<0.001
Admission quarter	Dec- Feb	1.00			<0.001
	Mar- May	0.68	0.69	0.59-0.81	
	Jun- Aug	0.76	0.79	0.68-0.92	
	Sep-Nov	0.74	0.80	0.69-0.94	
Presenting overnight (1700-0800)		1.81	1.75	1.56-1.96	<0.001
Medicine					
Baseline variables		OR	Adj OR	95% CI	p
Age	18-40	0.10	0.10	0.08-0.12	<0.001
	41-60	0.41	0.46	0.43-0.50	
	61-75	1.00			
	76-85	1.51	1.49	1.41-1.58	

	86+	2.28	2.25	2.12-2.38	
NEWS2 >5		6.18	5.01	4.78-5.23	<0.001
Sex (Females versus Male)		0.78	0.82	0.79-0.86	<0.001
Admissions preceding 30 days		1.62	1.56	1.49-1.64	<0.001
Admission quarter	Dec- Feb	1.00			<0.001
	Mar- May	0.83	0.89	0.84-0.95	
	Jun- Aug	0.79	0.89	0.84-0.95	
	Sep-Nov	0.83	0.94	0.89-1.00	
Admission year	2016	1.00			0.046
	2017	1.04	0.97	0.91-1.03	
	2018	0.93	0.93	0.87-0.98	
	2019	0.82	0.93	0.88-0.99	

3.3.3 Observations and early warning scores

Over the four years of the study, total observations recorded increased by 14% from 1,976,872 to 2,249,118 as shown in Table 3.4 below, with median observations per patient per day rising from 3 to 4 (Figure 3.2). If time taken to record observations is assumed to be 3 minutes 45 seconds [7], this equates to an increase of 85,000 minutes a month. 65% of observations were attributable to patients under medical specialties, 34% to patients under surgical specialties and 1% to patients discharged by the emergency department. The median admission NEWS2 remained stable at a score of one.

Figure 3-2 Median number of observations per patient per day with interquartile range

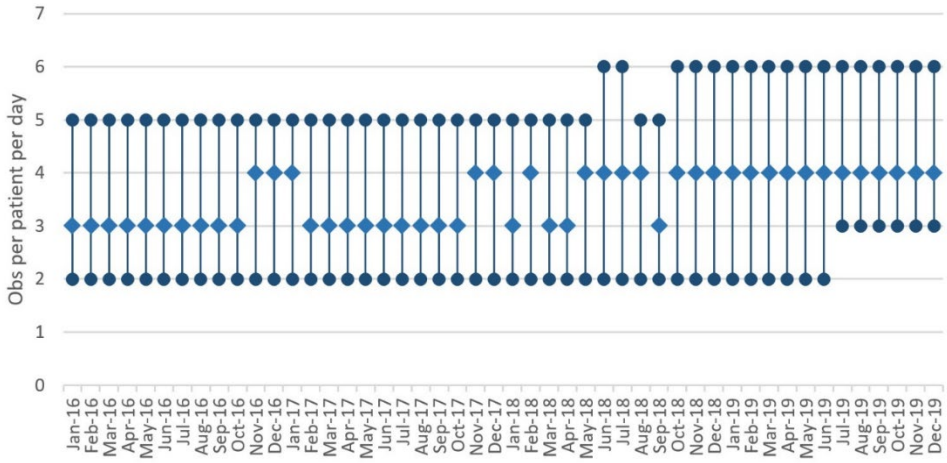


Table 3-4 Patterns of early warning score by specialty group and year

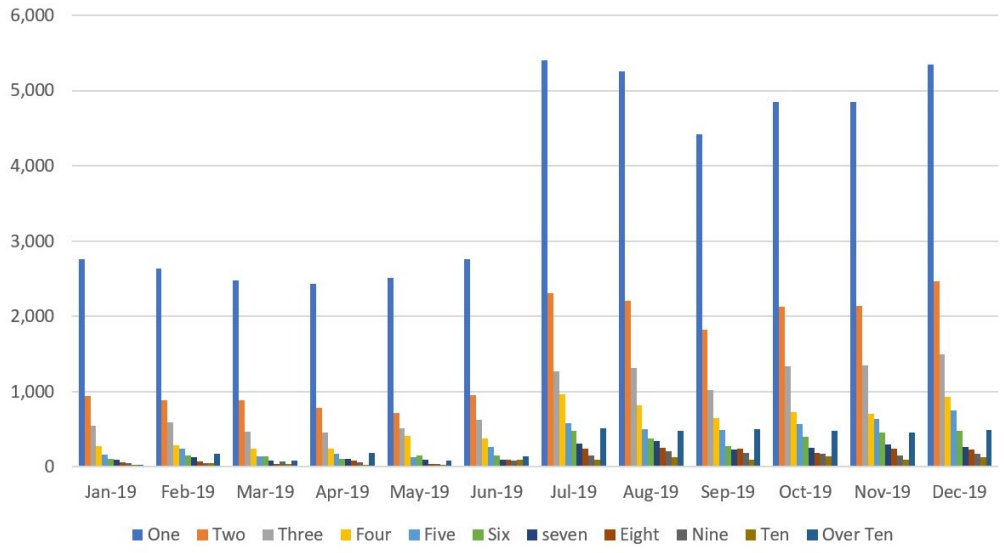
		2016	2017	2018	2019
Number of observations	Total	1,976,872	1,995,823	2,067,015	2,249,118
	Surgery	627,359	651,865	672,519	720,919
	Medicine	1,345,812	1,337,457	1,388,273	1,515,547
Median observations per patient per day(IQR)	Total	3 (2-5)	3 (2-5)	4 (2-5)	4 (2-6)
	Surgery	4 (2-5)	4 (2-5)	4 (2-6)	4 (3-6)
	Medicine	3 (2-5)	3 (2-5)	3 (2-5)	4 (3-6)
Median admission NEWS2 (IQR)	Total	1 (0-2)	1 (0-2)	1 (0-2)	1 (0-2)
	Surgery	1(0-2)	1(0-2)	1(0-2)	1(0-1)
	Medicine	1(0-3)	1(0-3)	1(0-3)	1(0-3)
Median NEWS2 (IQR)	Total	1(0-2)	2(0-2)	1(0-2)	1(0-2)
	Surgery	1 (0-2)	1(0-2)	1(0-2)	1(0-1)
	Medicine	1(0-3)	1(0-3)	1(0-3)	1(0-3)
% Observations followed by death in 24 hours	Total	0.48	0.50	0.46	0.40
	Surgery	0.20	0.22	0.17	0.15
	Medicine	0.59	0.63	0.60	0.51
Sensitivity for death in 24 hours NEWS2 5	Total	0.81	0.81	0.79	0.74
	Surgery	0.79	0.78	0.74	0.64
	Medicine	0.82	0.81	0.80	0.76
Specificity* for death in 24 hours NEWS2 5	Total	0.89	0.89	0.89	0.92
	Surgery	0.93	0.93	0.93	0.95
	Medicine	0.87	0.87	0.87	0.90
Sensitivity for death in 24 hours NEWS2 7	Total	0.62	0.62	0.60	0.57
	Surgery	0.58	0.61	0.53	0.47
	Medicine	0.63	0.63	0.61	0.58
Specificity for death in 24 hours NEWS2 7	Total	0.96	0.96	0.96	0.97
	Surgery	0.98	0.98	0.98	0.99
	Medicine	0.95	0.95	0.95	0.96
Area under ROC curve for death in 24 hours of NEWS2 (95%CI)	Total	0.921 (0.918-0.924)	0.918 (0.915-0.921)	0.909 (0.906-0.913)	0.910 (0.907-0.914)
	Surgery	0.928 (0.920-0.936)	0.920 (0.912-0.929)	0.901 (0.890-0.913)	0.891 (0.878-0.903)
	Medicine	0.915 (0.912-0.919)	0.913 (0.909-0.916)	0.904 (0.900-0.908)	0.908 (0.904-0.912)

* Specificity here refers to the percentage of observations not followed by death in 24 hours which fell below the threshold for escalation.

3.3.4 Workload

Recorded escalations to the medical registrar were relatively stable between January 2016 and June 2019. However, there was an increase of approximately 300% following the change from LEWS to NEWS2 in June 2019 (see Figure 2.4 and Table 2.4), when registrar escalations rose approximately 932 a month to over 3000 a month. This could mean an estimated additional 172-2068 hours a month depending on whether a five-minute review of the observations and notes or a full assessment and management plan, taking an estimated 60 minutes, is required. Upon reviewing the number of times that a patient was escalated to the registrar within a 24-hour period, an increase was seen across the spectrum. Patients escalated once in a 24-hour period rose from 2500 a month before the introduction of NEWS2 to 5000 a month after the introduction of NEWS2 and its associated escalation protocol. At the other end of the range, patients escalated more than 10 times in a 24-hour period rose from an average of 110 a month to an average of 486 a month following the introduction of NEWS2 and its escalated protocol (Figure 3.3).

Figure 3-3 Frequency at which a patient was escalated to the registrar in a 24 hour period by month 2019



This rise in recorded escalations is not reflected in patterns of predicted escalations calculated from retrospective analysis of LEWS and NEWS2 when each is looked at independently over the four years of the study. Both scores show a trend towards decrease in vital signs observation scores reaching the respective threshold score for registrar escalation (Figure 3-4). When examining percentage of scores above cut point

that were escalated, using the score in use at the time the observation set was recorded, just over 40% of observations with LEWS of 6 or more were escalated to the registrar while more than 60% of observations with a NEWS2 of 7 or more were escalated to the registrar.

Figure 3-4 Pattern of escalation to registrar plotted scores reaching LEWS escalation threshold by month and year

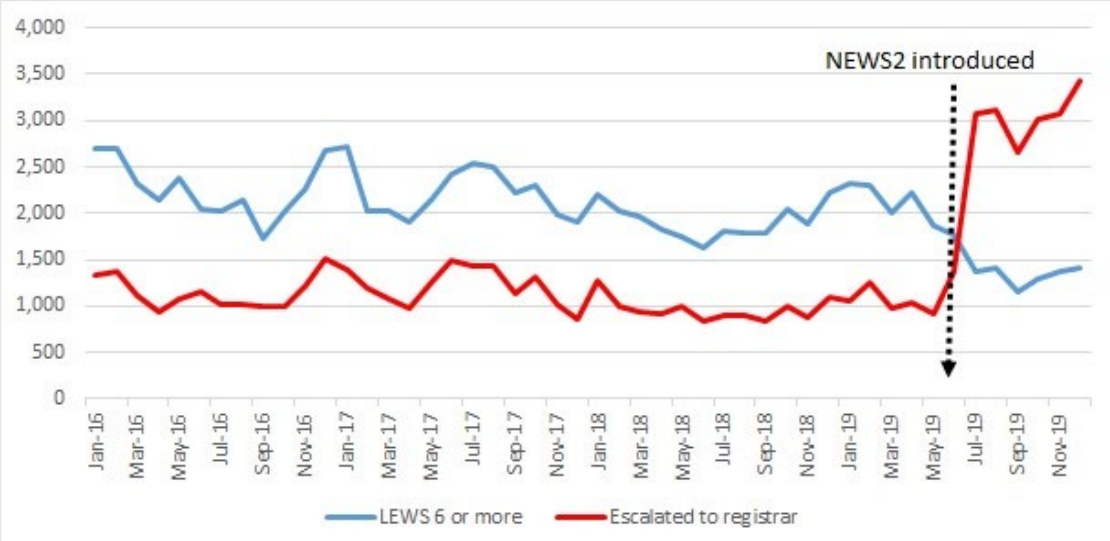


Table 3-5 Predicted escalations by scores reaching threshold and actual recorded escalations by year

		2016	2017	2018	2019**
Actual recorded registrar escalations *	Total	13,468 (0.7)	14,399 (0.7)	11,420 (0.6)	24,577 (1.1)
	Surgery	3,161 (0.5)	3,410 (0.5)	2,658 (0.4)	5,173 (0.7)
	Medicine	10,304 (0.8)	10,988 (0.8)	8,754 (0.6)	19,376 (1.3)
Median recorded registrar escalations per day*	Total	14 (6-23)	13 (5-22)	11 (5-18)	16 (7-33)
	Surgery	5 (3-9)	5 (3-10)	4 (2-8)	6 (3-11)
	Medicine	13 (5-23)	12 (5-22)	10 (4-17)	15 (7-30)
NEWS2 scores reaching threshold of 7(% total)	Total	80,505 (4.3)	83,732 (4.3)	80,438 (3.9)	63,085 (3.0)
	Surgery	13,757 (2.3)	14,188 (2.2)	11,917 (1.8)	9,680 (1.4)
	Medicine	66,731 (5.2)	69,515 (5.3)	68,504 (5.0)	53,335 (3.8)
LEWS scores reaching threshold of 6 (% total)	Total	26,484 (1.3)	26,116 (1.3)	22,537 (1.1)	20,144 (0.9)
	Surgery	4,732 (0.8)	4,545 (0.7)	3,436 (0.5)	3,192 (0.4)
	Medicine	21,734 (1.6)	21,566 (1.6)	19,091 (1.4)	16,937 (1.1)_
Number needed to evaluate for outcome of death in 24 hours NEWS2 score of 7	Total	15.4	14.2	15.1	14.5
	Surgery	20.2	16.8	21.8	20.8
	Medicine	14.7	13.8	14.3	13.8
Number needed to evaluate for outcome of death in 24 hours LEWS 6	Total	7.9	7.0	7.1	8.2
	Surgery	9.6	8.0	9.4	11.6
	Medicine	7.6	6.8	6.8	7.8
Actual number needed to evaluate from recorded escalations to registrar	Total	10.3	9.4	10.9	12.4
	Surgery	9.3	8.9	9.8	16.3
	Medicine	10.7	9.5	11.2	11.7

*NEWS2 introduced to NUHT in June 2019- before this point calculated retrospectively.

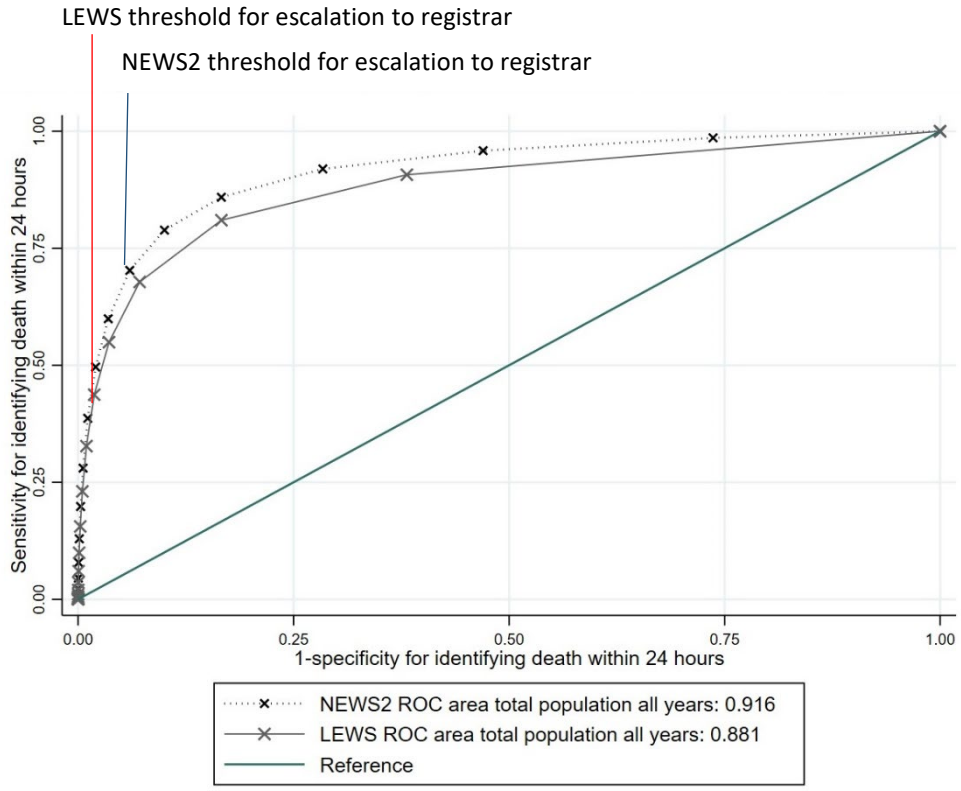
**Following introduction of NEWS2, LEWS high scores may be over-estimated as AVPU recording changed.

When using NEWS2 retrospectively to calculate number needed to evaluate (NNE), for every outcome of death within 24 hours detected at a threshold cut point of 7, 14.2-15.4

observations sets met the threshold for escalation (compared to 14.7-15.6 when applying NEWS1). The NNE for surgical patients at a threshold score of 7 was 16.8-21.8 observation sets for every death detected within 24 hours. In medical patients 13.8-14.7 observation sets met the threshold for escalation for every death within 24 hours. LEWS had a lower NNE at the threshold for registrar escalation. However, actual escalations did not match the predicted number of observations reaching threshold for escalation using either the LEWS prior to June 2019 or NEWS2 after June 2019 (Table 3-5) as a proportion of escalations were stopped by the clinical team if felt to be unnecessary or inappropriate.

On further analysis of performance. NEWS2 was found to have a larger area under the receiver operating characteristic curve than LEWS, but with cut points set for a far lower specificity (Figure 3-5).

Figure 3-5 Receiver operating characteristic curves for LEWS and NEWS2 for predicting death within 24 hours of an observation set in the total population



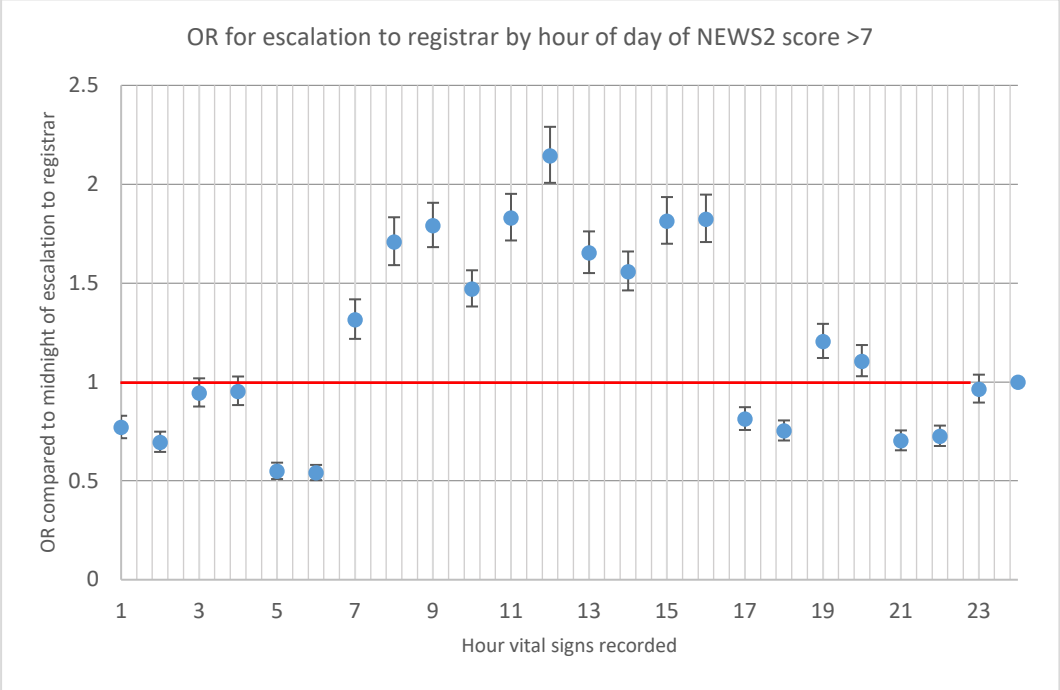
The area under receiver operating characteristic curve for NEWS2 in predicting outcome of death within 24 hours was similar between the two patient populations (0.910. 95%CI 0.908-0.911 in medicine and 0.912, 95%CI 0.907-0.917 in surgery). These values are comparable with similar study populations [8] and with the original NEWS1 protocol

(0.911 95%CI 0.090-0.913 in medicine and 0.919, 95%CI 0.914-0.924 in surgery). There was a significant reduction in area under ROC curve over time suggesting that by this measure NEWS2 was less able to predict which observations would be followed by death in 24 hours in 2019 than in 2016.

3.3.4.1 In versus out of hours escalation patterns

Out of hours observations comprised 75% of the total observations recorded. However, higher scores, i.e those over the threshold for escalation to the registrar using whichever system was in place at the time, were more likely to result in escalation to the registrar during the day than overnight. When the entire out of hours period was taken together the odds ratio for threshold scores being escalated to the registrar out of hours versus in hours as defined previously was 0.33 (95%CI 0.32-0.34). The odds ratios for escalation to the registrar by hour of the day of vital signs observation sets scoring at or above the NEWS2 threshold can be seen in figure 3-6.

Figure 3-6 OR for escalation of vital signs scoring at or above the threshold for registrar review as per NEWS2 protocol (Midnight as baseline)



3.3.5 National use of NEWS2 and application of escalation protocols

74 Trusts covering over 100 hospitals across England responded to the freedom of information request sent out in March 2022. 65(%) out of the 74 Trusts employed electronic observations platforms (see Figure 3-7 for distribution of trusts responding) to deploy NEWS2 and its associated escalation protocol, with 24 different platforms reported as being in use. Two further Trusts indicated they were looking to deploy electronic observations in the future. 12 of these Trusts reported employing automatic escalation of observations to the registrar. The cut points reported for escalation varied, with reported thresholds scores of 4-5 to more junior doctors, 5 or 7 to the registrar and 5 or 7 to critical care outreach teams. Trusts not deploying automated escalation reported relying on a combination of nursing staff escalation based on advised actions at set scores, and dashboards displaying threshold scores across the hospital to highlight high scores. One Trust reported using an additional risk assessment based on highest score in the preceding 12 hours alongside current NEWS score to assist in clinical judgement.

Figure 3-7 Geographical distribution of trusts responding to freedom of information request regarding use of electronic observations, early warning scores and software platforms

3.4 Discussion

3.4.1 The Hospital Population

Our data describe the characteristics of the inpatient population and trends over the study period. They also highlight the impact of introducing a new early warning score to an stablished electronic observations and escalations system. The observed increase in the number of admissions year on year, with a smaller increase in bed days, associated with a trend towards a decrease in length of stay is consistent with figures reported by the King's Fund into NHS activity which described a reduction of 3000 beds across the NHS from 2016-2019 [9]. Over the same period inpatient elective and emergency attendances rose by 9% nationally compared with 19% in this dataset [10]. Mortality reduced from 3.7 to 3.1% in the overall hospital inpatient population between 2016 and 2019. This fall in mortality is consistent with overall patterns of mortality in Nottinghamshire between 2016-2019 and is not offset by a higher proportion of deaths in the community [11]. This pattern of a higher frequency of shorter admissions was associated with an increase in the number of vital signs observations being collected per day.

Differences in patient characteristics were also identified between different inpatient populations. Patients admitted under surgical specialties accounted for 37% of admissions, 32% of observations, but only 16% of scores over seven and 12% of deaths. Medical inpatients had a longer length of stay, higher mortality and higher number of admissions.

In analysing the factors associated with length of stay and mortality, several were common to both medical and surgical inpatients. Advancing age, high NEWS score at admission and time of year were risk factors for increased mortality and longer length of stay in all populations. However, there were differences in terms of the impact of gender, readmission, and time of day admitted. For example, being female was associated with a higher mortality risk in surgery[15], potentially contributed to by the inclusion of cardiac surgery patients [16], but a lower risk in medicine. Admission overnight, between 5pm and 8am, was associated with an increased mortality risk in surgery but not medicine when other factors were adjusted for in the multivariate model. A discharge within the 30 days preceding the current admission was associated with a shorter than median length of stay in surgery and a longer than median length of stay in medicine.

The data on the mortality trend in the second half of 2019 was in keeping with the downward trend seen in previous years. However, multiple new ways of working have been introduced into hospitals over the years, including focusing on falls, pressure care, sepsis 6, early consultant review etc; consequently, it is very difficult to establish which of these may or may not have affected mortality. In addition, having only a few months of trend after the introduction of NEWS2, due to the emergence of COVID-19 in 2020, means it is not possible to distinguish any effect of the new score from seasonal fluctuations in disease.

3.4.2 Prognostic ability of NEWS2

The prognostic ability of NEWS2, as measured by area under ROC curve for outcome of death within 24 hours, reduced between 2016 and 2019. One contributory factor could be a change in the way conscious level was recorded on the Nervecentre platform. Analysis of vital signs patterns before and after the change in early warning score to NEWS2 demonstrated a drop in observations coded as having reduced conscious level according to ACVPU (a tool that rates conscious level based on whether someone is Alert, has new Confusion, is responsive to Voice or Pain or is Unresponsive), following the introduction of NEWS2. The reduction in mortality over time may also have influenced the decline in the performance of NEWS2.

In terms of performance in different inpatient populations, only one previous study has examined the prognostic ability of NEWS in predicting outcome in medical versus surgical inpatients [8]. The proportion of observations followed by death within 24 hours was 0.21% in surgery and 0.69% in medicine, comparable to our study. The primary method for judging score performance in these populations was area under receiver operating characteristic curve for outcome of death, ICU admission, cardiac arrest and combined within 24 hours of an observation. By this measure the performance of NEWS2 was not significantly different between the two groups and performed at least as well in surgical patients as medical patients, a result replicated in this study. The authors also used a measure of workload and detection (sensitivity), both of which clearly showed a difference between the populations. Again, results were comparable to this study, despite their use of combined outcomes to report this metric in comparison to our use of death within 24 hours as an outcome. This supports the view that these two cohorts represent distinct populations with different characteristics requiring different management.

As with all studies analysing vital signs linked to outcome, it is only possible to see the association with scores and that outcome. It is not possible to determine where a high score has triggered an intervention that averts an outcome as intended, or where factors known to impact outcome such as staffing [18] are not available. In order to establish a causative link between use of NEWS2 and mortality a randomised control trial would be needed. The use of death within 24 hours as an outcome, rather than cardiac arrest, ICU admission or combined, means this study cannot be compared directly with the only previous study examining NEWS in surgical and medical populations [8].

3.4.3 The Impact of introducing a new Early Warning Score on the system

The two different early warning scores, LEWS and NEWS2, had varying effects on demand as defined by recorded escalations to the on call team or the registrar. Despite both scores showing downwards trends in observations reaching the threshold for escalation over the course of the study, the recorded escalations to the registrar more than trebled partway through 2019. Rising from an average of 932 a month in the 6 months July to December 2018 to an average of 3,062 a month in the 6 months July-December 2019, when NEWS2 was introduced. It is not possible to match the cut points of the two scores as the shapes of the receiver operating curves means they do not overlap except at extremes of sensitivity and specificity (Figure 2.5), where any cut point would be meaningless. However, NEWS2 has a higher sensitivity and lower specificity than LEWS at the escalation thresholds with equivalent actions, resulting in a higher number of escalations to the registrars, including a rise in patients being escalated multiple times in a 24-hour period (Figure 2.3). In addition to the statistical performance, the human factors element of introducing a new early warning score should also be considered. A higher proportion of observations reaching threshold score were escalated to the registrar following the introduction of NEWS2 than had been the case with the previous local score. One explanation for this is familiarity with the score. It is possible that given lack of experience with a new score staff felt less able to use their own judgement where it contradicted the protocol. It is also possible that the Hawthorne effect played a role due to the increased monitoring that was carried out in the months after NEWS2 and its associated protocol was introduced.

It could be argued that an increase in NNE, far from reflecting a failure to recognise instability points to a reduction in adverse outcomes as more patients were reviewed by senior staff. However, there is insufficient data to support this theory and the increase in

escalations is problematic from a workforce perspective as there has been no associated increase in registrar numbers in training. Without changes to either workforce numbers or NEWS2 escalation thresholds, increased demand inevitably contributes to a delay in clinical review for some patients or an impact on the ability of registrars to complete other aspects of their job important to training and development.

This conclusion is supported by theories relating to mental workload which suggest that there is a limit to the amount of processing that can be carried out at a particular time [105], and that significant levels of multitasking may lead to increased psychophysical strain. The need for data-driven rotas which account for differences in demand by specialty or time of day has previously been mooted as a way to deal with this [106]. Another response is through greater empowerment of nursing staff in terms of their assessment of patient condition [12-14]. Or through the use of specialist nurses acting as a first point of call for deteriorating patients and liaising with critical care outreach teams (CCOT). Studies examining the impact of introducing CCOT into a system for responding to deteriorating patients found improvement in staff in decision making and early access to ICU [12]. However, robust evidence remains lacking [13] and CCOT comes at an increased cost.

3.4.4 How hospitals are using NEWS2

These differences in outcome could be used influence the composition and application of early warning scoring systems in terms of thresholds for escalation; however, any difference in the applicability of NEWS2 has to be balanced against the benefits of having a single standardised score for any deteriorating adult patient in terms of familiarity with score and benchmarking of care. Moreover, monitoring of patients is reliant not just on the score used, but how the absolute score, and any need for clinical review, is communicated to medical staff and the clinical response to it. The Freedom of Information response shows significant variation in how NEWS2 has been adopted across the NHS. This includes the use of different threshold cut points in different trusts, a differential response to the threshold scores and varying staffing responses in term of seniority/experience. It is apparent that despite a single mandated national scoring system the response to a deteriorating patient is still varied. This could suggest that, as with the chronic respiratory scale many hospitals developed alongside the first NEWS, NHS hospitals are finding ways to use NEWS2 that are compatible with their system and staffing resources. This highlights the fact that although a single system has benefits,

there may need to be refinement regarding the cut points applied in the real world and, ideally, prospective studies to refine implementation. This would confirm whether specific cut points for patients in different specialty areas, for example medical versus surgical specialties, would be beneficial as our data, and others [6] suggests.

There is also disparity in response to a patient identified as at risk of deterioration in the deployment and make up of rapid response teams, (acute response teams- RRT/ medical emergency teams- MET). These generally consist of a number of on call doctors including those with critical care or airway skills. The MERIT study was a cluster randomised control trial of the introduction of MET to 23 hospitals. It reported that despite a higher number of emergency referrals, the introduction of a MET did not lead to a reduction in mortality [17], although this may reflect length of follow up, as a further study reported lower mortality after a longer period and a change in team composition [12, 13].

3.4.5 Conclusion

In conclusion, our study illustrates clear differences in population characteristics and mortality between patients admitted under medical and surgical specialties and an associated difference in ability of NEWS2 to predict outcome at the current protocol thresholds. The increase in escalations following switch to NEWS2 also highlights the potential workload impact of changes to scores and associated escalation protocols.

4 Examining the performance of NEWS2 in patients with respiratory disease and exploring use of NEWS2 patterns in improving prognostic ability.

4.1 Introduction

4.1.1 Early Warning Scores in Respiratory Medicine

The preceding chapter described the adult hospital population captured within the database being analysed within this thesis, with some initial analysis of the differences between the medical and surgical inpatient populations with relation to mortality and length of stay. The study described in this chapter builds on this to examine concerns raised regarding the performance of NEWS2 in the setting of chronic disease where baseline vital sign values can differ from those seen in the population from which NEWS was derived, and where physiology can react differently to the stresses of acute pathology [58, 107].

An important example is patients with Chronic Obstructive Pulmonary Disease (COPD), the third leading cause of death worldwide. COPD is an airways disease characterised by chronic changes in respiratory rate, oxygen saturation, heart rate and carbon dioxide retention. Altered baseline physiology may elevate the NEWS2 score, leading to unnecessary medical interventions in stable patients, alert fatigue in medical staff (reducing clinical response to a high scoring patient[35]), inappropriate oxygen use, or misplaced clinical reassurance in an unstable patient [108].

4.1.2 Aims of Study

We set out to determine the i) the statistical discrimination of NEWS2 in patients with respiratory disease; ii) the impact of COPD on the ability of NEWS2 to predict risk of adverse outcome; and iii) whether the performance of the NEWS2 could be improved, both in terms of detecting deterioration and de-escalating care, in patients with respiratory disease.

4.2 Methods

4.2.1 Source of data

This study was carried out using the database described in Chapter 2. For the purposes of this study, patients admitted under the care of Respiratory Medicine were isolated for analysis. These data were then split into an initial derivation cohort from April 2015 to March 2017, and a validation cohort from April 2017 to March 2019.

4.2.2 Participants

All admissions in patients aged 18 years or older, completed within the study period admitted to and discharged from Respiratory Medicine were included. Any vital signs coded as 'End of Life Care' (meaning all interventions would be aimed at palliation of symptoms rather than prolonging life) were removed from the analysis.

4.2.3 Data Handling

NEWS2 oxygen saturation scale 1 was applied to all patients without a diagnosis of COPD with target saturations of 94-98%. Scale 2 (which adjusts for patients at risk of hypercapnic respiratory failure) was applied to all patients with a diagnosis of COPD in line with previous research [77], identified by presence of an ICD10 code for COPD at any point during that admission.

Current NEWS2 score was applied as an independent variable and as part of novel bivariate models combining current NEWS2 score with the pattern of NEWS2 score, both over the preceding 24 hours and throughout admission, to assess ability to predict death within 24 hours of an observation. Death was used as the outcome rather than ICU admission as several factors influence ICU admission (bed availability, staffing etc), not just clinical status.

Initially, 4 hour time points were created throughout a patient's admission in order to address possible bias created by a patient with a higher score having more frequent observations. However, in this patient population, 25% of observations were taken at an interval of more than 7 hours and therefore this approach would mean carrying forward too many observations from the preceding period to allow a robust analysis. Therefore all observations were applied as independent points in line with previous early warning score research.

Several scoring patterns were generated to explore whether patterns in NEWS2 score could be incorporated with current score to improve prognostic ability. These included first NEWS2 score, early patterns in rise and fall, difference between current and previous NEWS2 value, labelled as delta NEWS2, moving average of preceding 5 scores, maximum value, minimum value, standard deviation of scores, and mean of scores over the preceding 24 hours and since admission. The patterns were used to create restricted cubic spline models with three knots at the placement recommended by Harrell [109]. Univariate models were created using the uvrs package in STATA. Each variable was then combined with current NEWS2 score using the mvrs package to create bivariate restricted cubic spline models. As an additional analysis, a predictive additive model was created using maximum NEWS2 score in the preceding 24 hours and current NEWS2 score.

Ability to predict death was assessed using several different approaches. Sensitivity and specificity at the clinical cut points of 5, 5 or a single vital sign score of 3, and 7 were calculated to reflect the clinical application of the score. NEWS2 was also treated as a continuous ordinal and evaluated using area under receiver operating characteristic curve (ROC curve), a plot of sensitivity over 1-specificity, and area under precision recall curve (PR curve), a plot of precision (positive predictive value) against recall (sensitivity) as appropriate in the whole population, and then in separate cohorts defined by COPD diagnosis. Use of area under the PR curve was used in addition to area under the ROC curve as the latter can be affected disproportionately by small improvements in prognostic ability in the setting of a data set with skewed outcomes, with a very small percentage of observations associated with adverse outcomes, as seen in hospital populations. As with area under ROC curve, the higher the area under the PR curve, the better the model performance.

Initial analysis was performed on admissions completed between 1st April 2015 and 31st March 2017, with models derived from this cohort. Analysis for validation of these models was then carried out on admissions completed between 1st April 2017 and 31st March 2019.

4.3 Results

4.3.1 Study Population

There were 7487 completed admissions from 5136 individual patients to the Nottingham University Hospitals Trust Respiratory Department during the initial two year study

derivation period from April 2015 to March 2017 (Figure 4-1) and 8739 admissions from 5928 individual patients during the second validation period from April 2017 to March 2019 (Figure 4-2). 218 admissions with only one set of recorded observations were removed from the derivation period and 254 admissions with only one set of recorded observations were removed from the validation period.

There were 5362 admissions in NEWS2 scale 1 cohort (no diagnosis of COPD) in the derivation cohort and 6351 in the validation cohort. The NEWS2 scale 2 cohort (with COPD) contained 2125 patients in the derivation cohort and 2381 patients in the validation cohort respectively. Admission demographics are detailed in Table 4-1**.** 249 sets of observations marked as 'End of Life Care' were excluded from analysis in the derivation cohort and 98 from the cohort.

Figure 4-1 Patients with respiratory disease completing admission between 1st April 2015 and 31st March 2017-derivation cohort

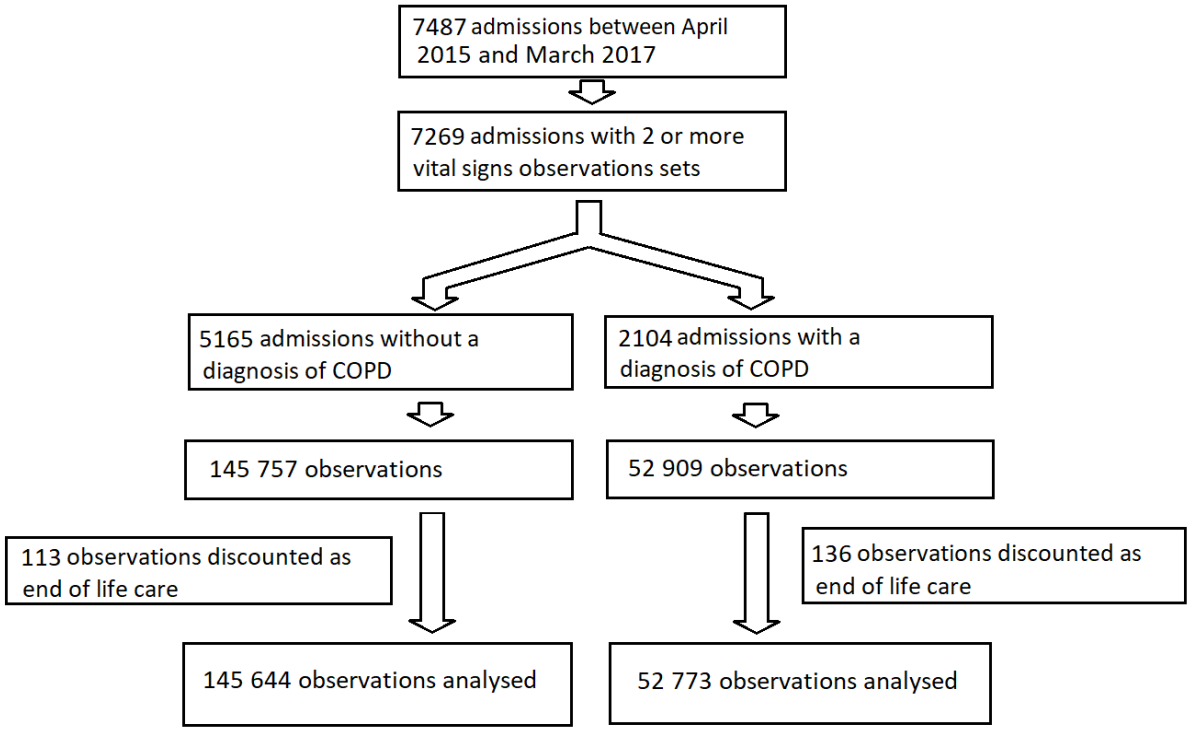


Figure 4-2 Patients with respiratory disease completing admission between 1st April 2017 and 31st March 2019- validation cohort

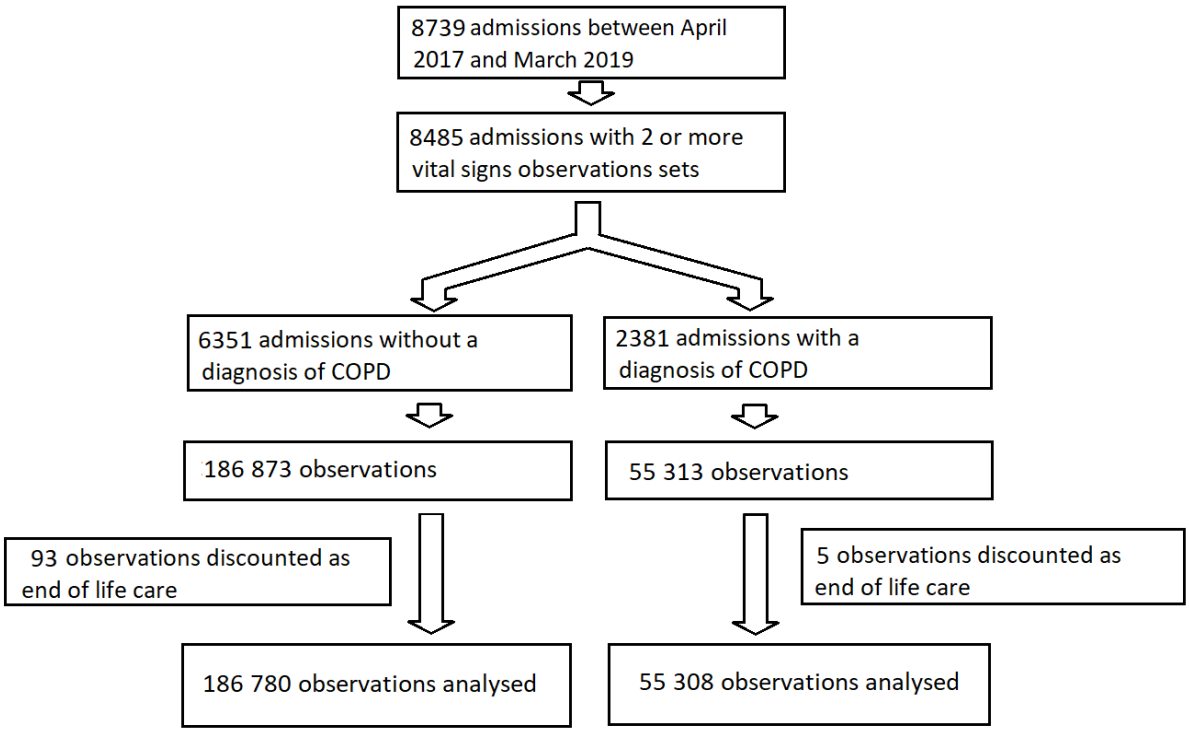


Table 4-1 Derivation and Validation cohort demographics

	April 2015- March 2017			**April 2017- March 2019**		
	Respiratory (total)	**Non- COPD (scale 1)**	**COPD (scale 2)**	**Respiratory (total)**	**Non- COPD (scale 1)**	**COPD (scale 2)**
Admissions (n)	7269	5165	2104	8485	6351	2381
Female (%)	3953 (54.4)	2775 (53.7)	1178 (56.0)	4718 (54.0)	3402 (53.5)	1316 (55.3)
Median age (IQR)	71 (61- 81)	71 (61- 81)	71 (61- 76)	71 (56-76)	66(51-76)	71 (61-76)
Median Length of Stay in days (IQR)	4 (2-8)	4 (2-8)	3(2-7)	3 (1-7)	3 (1-7)	3 (1-6)
In hospital mortality (%)	413 (5.7)	328 (6.4)	85 (4.0)	470 (5.5)	398 (6.5)	72 (3.1)

4.3.2 Overall performance of NEWS2 in total respiratory population

In the total respiratory population NEWS2 demonstrated a sensitivity of 0.87 and specificity of 0.72 at a cut point of 5 for predicting death within 24 hours of an observation set. Sensitivity increased to 0.89 where observations with a single vital sign scoring 3 were added to scores of 5 or more, at the expense of a reduction of specificity to 0.67. At a cut point of 7, sensitivity for predicting death within 24 hours was reduced to 0.68 and specificity increased to 0.90.

Area under the ROC curve for NEWS2 in the overall respiratory population was 0.888 (95% CI 0.881-0.895) in the derivation cohort of April 2015 to March 2017 and 0.880 (95% CI 0.873- 0.887) in the validation cohort of April 2017 to March 2019 for predicting death within 24 hours. Area under the PR curve was 0.140 in the derivation cohort and 0.133 in the validation cohort. Each point increase in NEWS2 score increased the odds ratio for death within 24 hours of an observation by 1.72 (95% CI 1.69-1.74) in the derivation cohort and 1.70 (95% CI 1.68 -1.72) in the validation cohort.

Workload

The additional clinical demand, a surrogate for workload (i.e. patient review by nurse or doctor), that high NEWS score led to can be seen in the number of observations reaching the threshold for review that were then not followed by an outcome within 24 hours. For example, 34 observations met the criteria for escalation and clinical review for every observation followed by death within 24 hours of that score at a cut point of 5, meaning there were 33 false positives to every true positive. This increased to 38 if observations scoring 3 in a single vital sign were included. 16 observations per outcome identified met the criteria for escalation at a cut point of 7. These values were similar to those seen in the validation cohort of patients admitted between April 2017 and March 2019, as demonstrated by the comparison of sensitivity, specificity and NNE at different cut points in Table 4-2 below .

4.3.3 Performance of NEWS2 in patients without a diagnosis of COPD applying oxygen target saturation scale 1

The prognostic ability of NEWS2 for predicting death within 24 hours was similar in the scale 1 cohort (no COPD) to the overall population. Sensitivity for predicting death within

24 hours at a cut point of 5 was 0.90 and specificity 0.70. Addition of observations where a single vital sign scored 3 again demonstrated a small increase in sensitivity to 0.91 and reduction in specificity to 0.64. A cut point of 7 demonstrated a sensitivity of 0.72 and specificity of 0.89 (Table 4-2 A projected 32 observations reached the threshold for clinical review/escalation per outcome at a cut point of 5, 37 per outcome if observations where a single vital sign scored 3 were included, and 16 per outcome at a cut point of 7 .

Area under ROC curve analysis for death within 24 hours of an observation was 0.895 (95% CI 0.887-0.903) and area under the PR curve was 0.146 in the derivation cohort. In the validation cohort, area under the ROC curve was 0.879 and area under the PR curve was 0.139.The odds ratio for death within 24 hours of an observation for each point increase in NEWS2 score was 1.72 (95% CI 1.69-1.75) in the derivation cohort and 1.70 (95%CI 1.67-1.72) in the validation cohort.

4.3.4 Performance of NEWS2 in patients with a diagnosis of COPD applying oxygen target saturation scale 2

Sensitivity at a cut point of 5 for outcome of death within 24 hours was reduced to 0.77 In the scale 2 cohort, with a higher specificity of 0.77 when compared to the scale 1 cohort. Adding in observations with scores of 3 in one vital sign increased sensitivity to 0.81 with specificity reduced to 0.74. For a cut point of 7 sensitivity was 0.53 and specificity was 0.93.

39 observations met the criteria for clinical review/escalation at a cut point of 5 per outcome identified of death within 24 hours. 41 observations per outcome identified met the criteria for escalation if observations containing a single vital sign scoring 3 were included and 17 observations at a cut point of 7.

Area under the ROC curve analysis was 0.857 (95% CI 0.838-0.877) and area under the PR curve was 0.114 in the derivation cohort (see Table 4-3) for comparison of area under ROC curve for NEWS2 score accuracy in predicting death within 24 hours in patients with and without a diagnosis of COPD). Area under ROC curve was 0.878 and are under PR curve was 0.100 in the validation cohort. The odds ratio per point increase in NEWS2 score was 1.70 (95% CI 1.65-1.76) in the derivation cohort and 1.76 (95%CI 1.70-1.83) in the validation cohort .

Figure 4-3 Area under the ROC curve graph for death within 24 hours of observation set- comparison of scale 1 cohort with no diagnosis of COPD and scale 2 cohort with a diagnosis of COPD recorded.

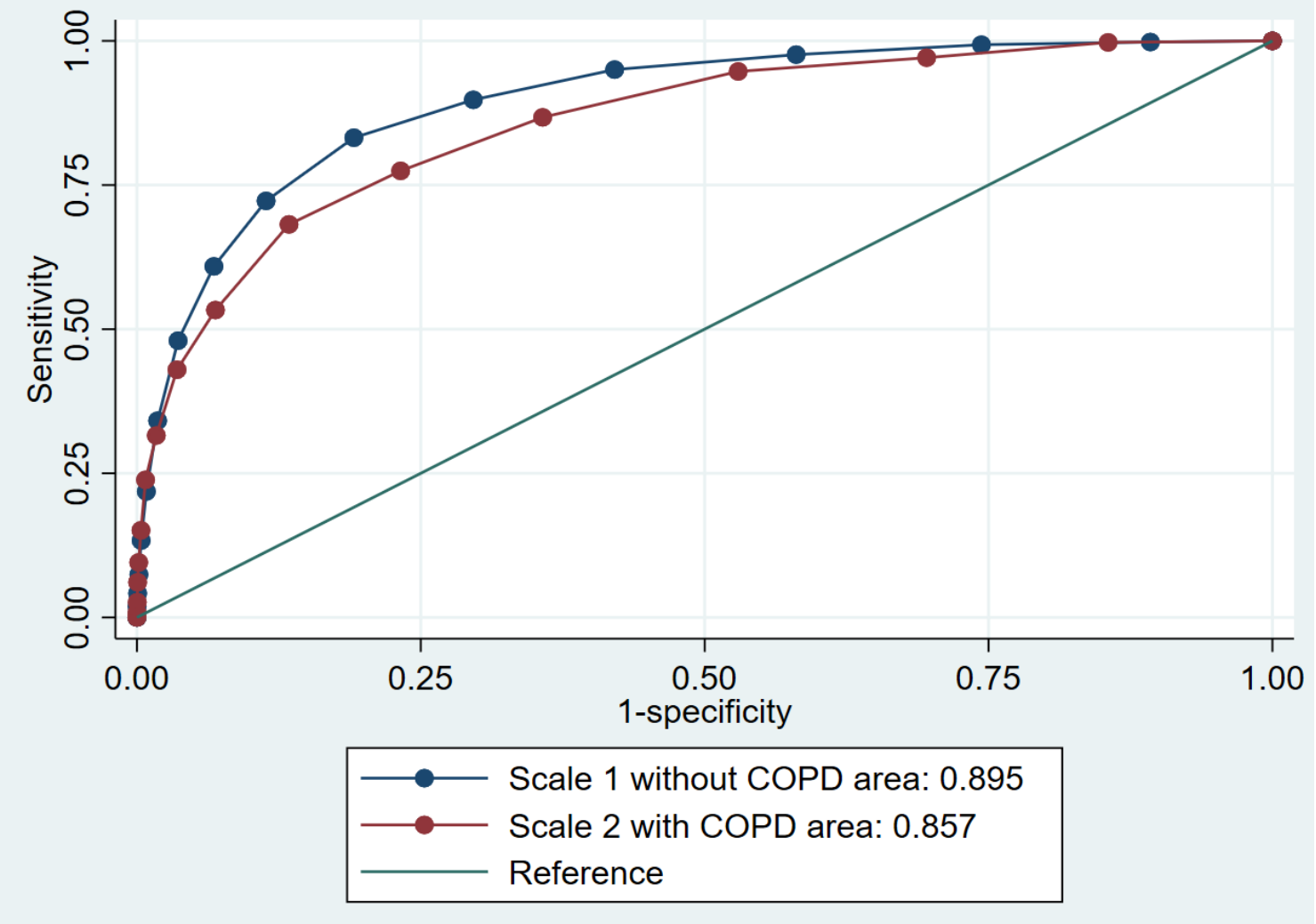


Table 4-2 Sensitivity, Specificity and NNE values of NEWS2 at cut points of 5 and 7 in their derivation and validation cohorts

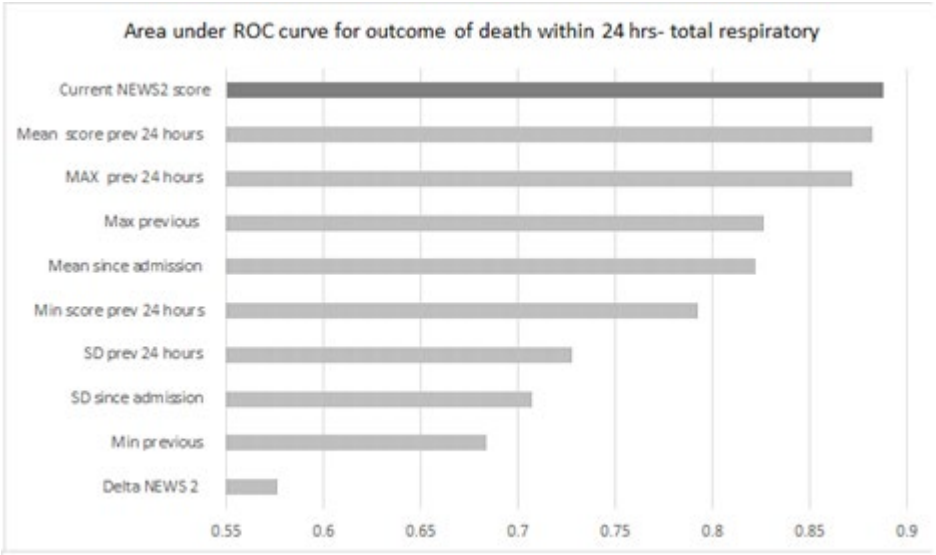
	Derivation cohort April 2015- March 2017			Validation cohort April 2017- March 2019		
Cut point 5	**Sensitivity**	**Specificity**	**NNE (95% CI)**	**Sensitivity**	**Specificity**	**NNE (95% CI)**
NEWS2 in total respiratory population	0.87	0.72	34.2 (32.5-35.7)	0.88	0.69	36.1 (34.7-37.8)
NEWS2 in patients without COPD (Scale 1)	0.90	0.70	32.4 (30.7-34.1)	0.90	0.66	35.2 (33.6-36.9)
NEWS2 in COPD (Scale 2)	0.77	0.77	42.4 (37.9-47.5)	0.81	0.79	41.8 (37.3-46.6)
Cut point 5 or single vital sign score of 3						
NEWS2 in total respiratory population	0.89	0.67	43.8 (41.8-45.9)	0.89	0.64	47.7 (45.7-49.8)
NEWS2 in patients without COPD (Scale 1)	0.91	0.64	36.7 (34.9-38.6)	0.90	0.60	41.8 (40.0-43.8)
NEWS2 in COPD (Scale 2)	0.81	0.74	71.7 (64.7-79.3)	0.85	0.76	79.1 (71.1-88.0)
Cut point 7						
NEWS2 in total respiratory population	0.68	0.90	16.2 (15.4-17.1)	0.70	0.88	17.8 (16.9-18.6)

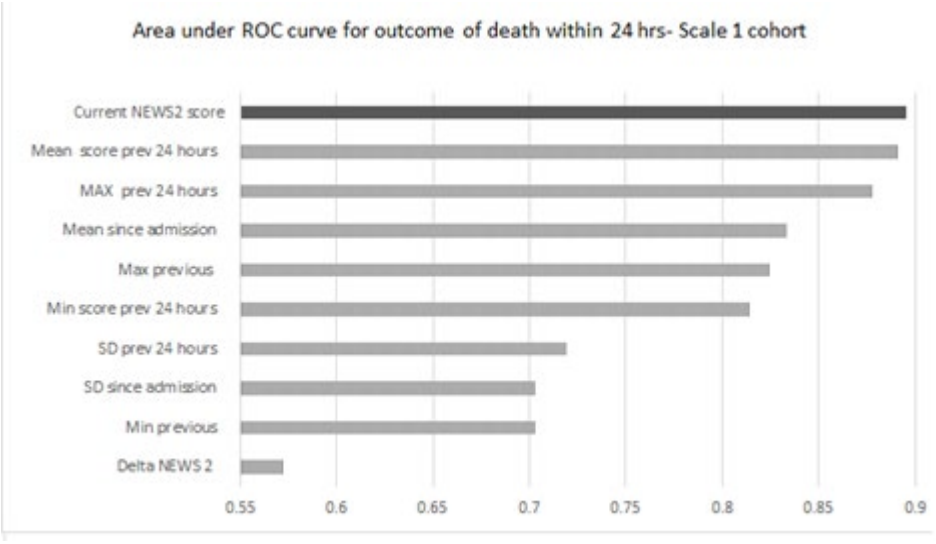
NEWS2 in patients without COPD (Scale 1)	0.72	0.89	16.0 (15.1-16.9)	0.72	0.87	17.8 (16.9-18.8)
NEWS2 in COPD (Scale2)	0.53	0.93	18.9 (16.5-21.7)	0.57	0.94	17.4 (15.3-19.8)

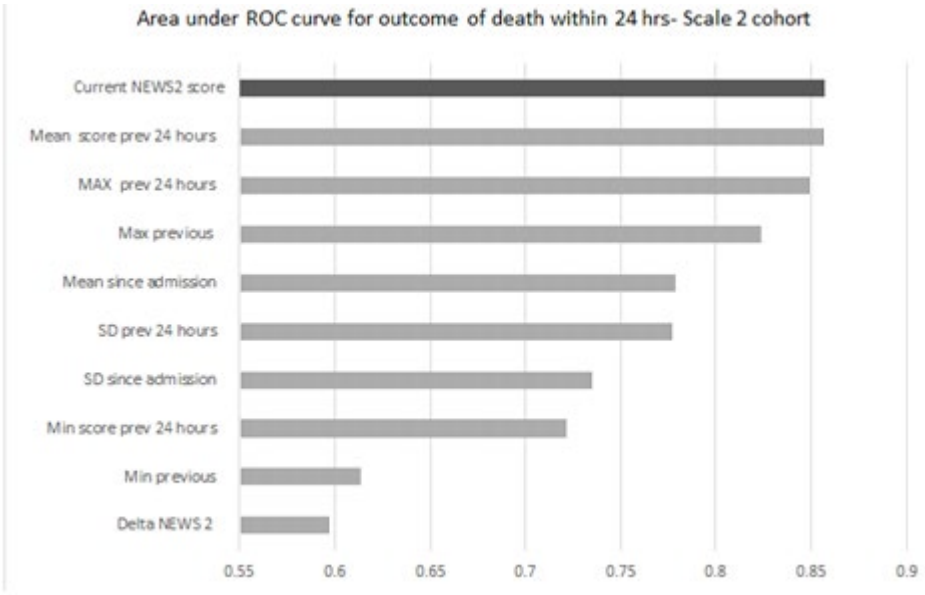
4.3.5 Pattern variables of NEWS2

Maximum and mean NEWS2 in the preceding 24 hours demonstrated similar area under ROC curve analysis to current NEWS2 for outcome of death in 24 hours, see

Figure 4-4 Comparison of area under ROC curves for univariate analysis of NEWS2 and other trend variables for outcome of death within 24 hours

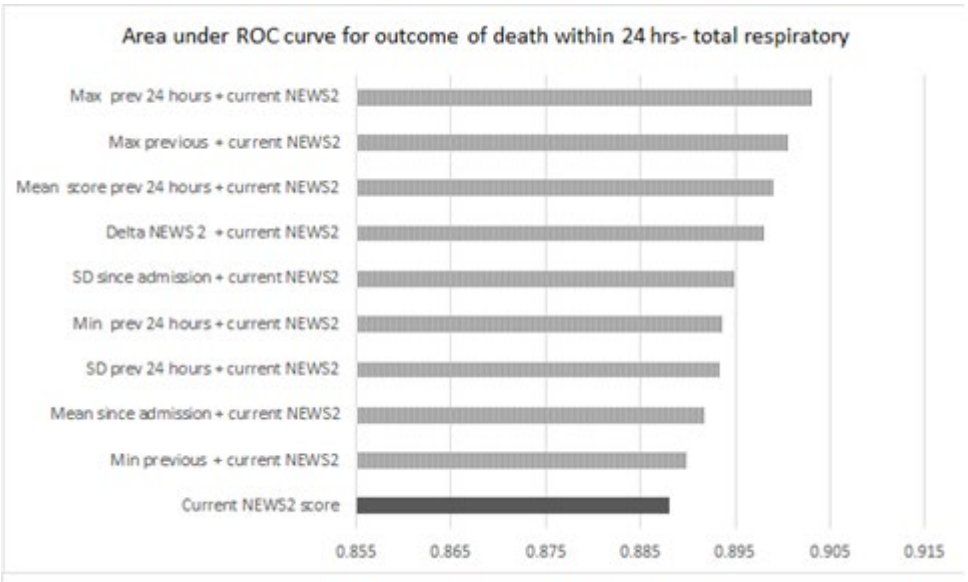


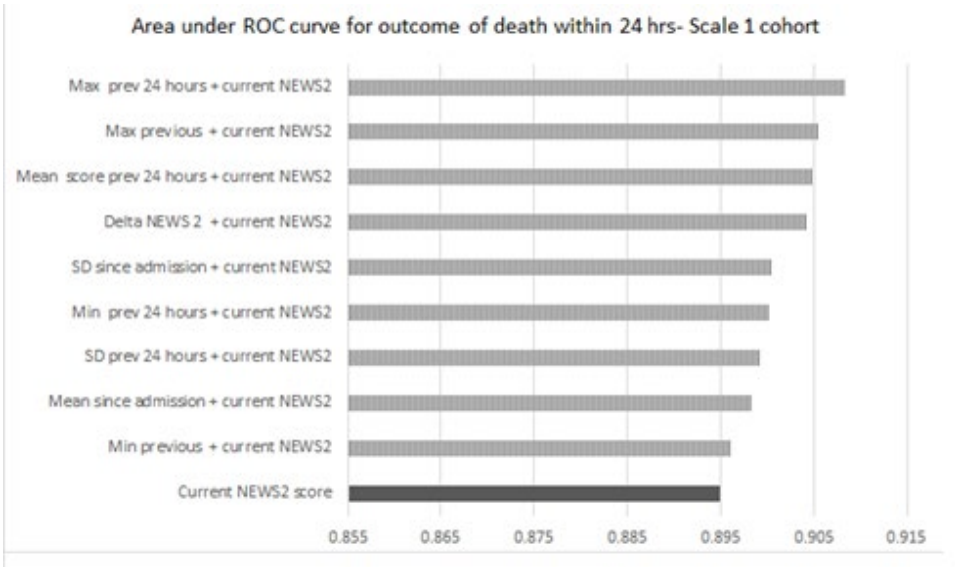


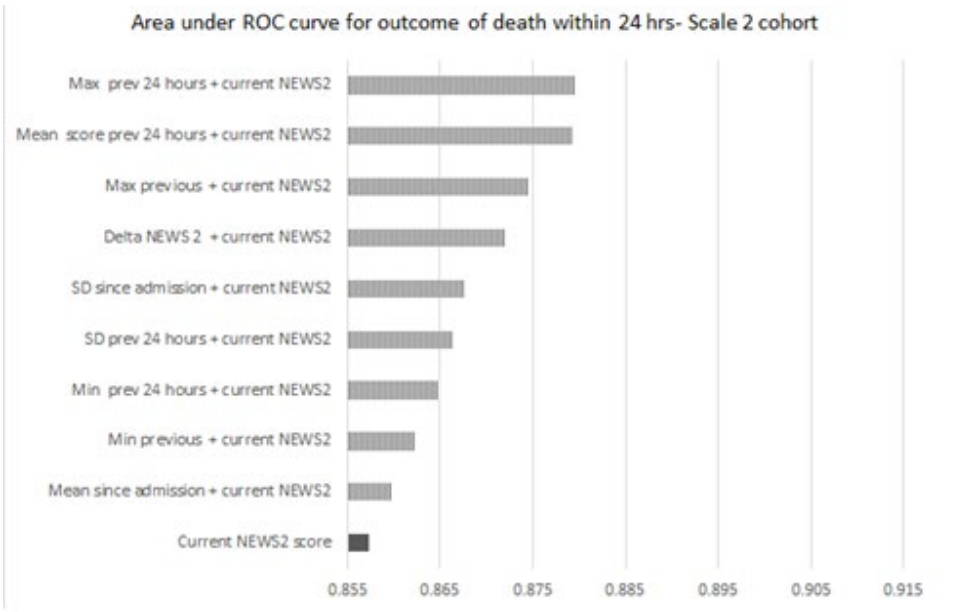


Here the 'SD' represents the standard deviation in NEWS2 scores for an individual over the entire admission or preceding 24 hours and delta NEWS represents the difference between each NEWS2 score and the previously recorded score. Restricted cubic splines were created as described in the methods from each trend variable and tested in terms of ability to predict death within 24 hours.

Figure 4-5 Comparison of area under ROC curves for bivariate analysis of NEWS2 combined with each trend variable in turn for outcome of death within 24 hours







Improvement in prognostic ability was seen in all bivariate restricted cubic spline models when compared to current NEWS2 alone (Figure 4-5). The model with highest prognostic ability for death within 24 hours combined maximum score in the preceding 24 hours with current score, with a ROC curve value of 0.903 in the total population, 0.908 in the Scale 1 cohort and 0.880 in the Scale 2 cohort.

An additive model created using maximum score in the preceding 24 hours and current score had equal prognostic ability to the spline model using the same components, with ROC curves for outcome of 0.902 in the overall population, 0.907 in the Scale 1 cohort and 0.880 in the Scale 2 cohort. This is also reflected in the area under precision recall (PR curves) curves shown in Table 4-3.

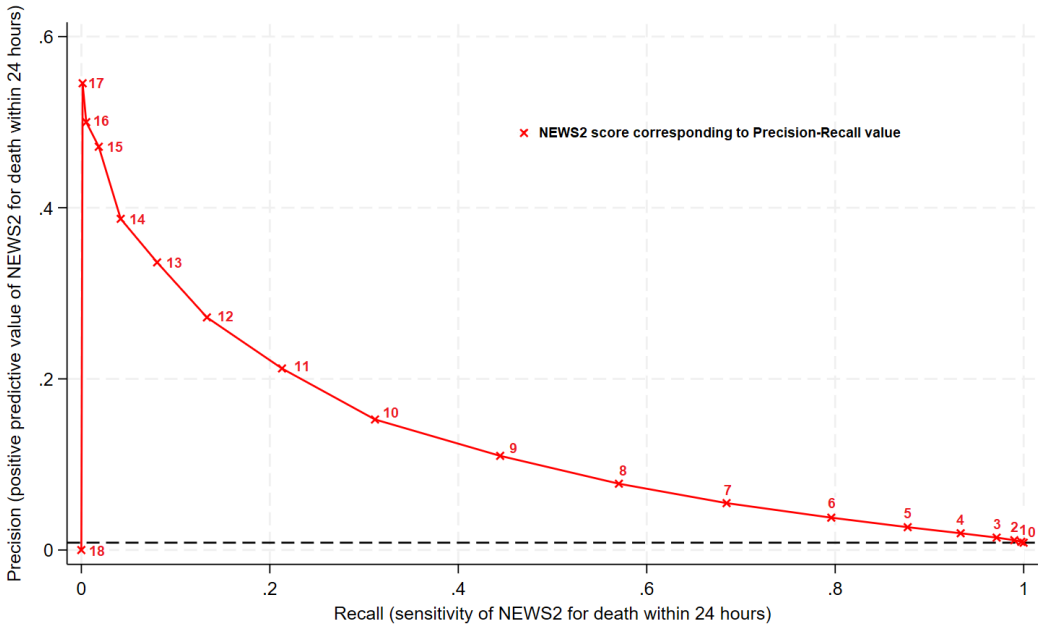
Table 4-3 Area under ROC and PRC for NEWS2 and additive score combining current NEWS2 and maximum NEWS2 in the preceding 24 hours

Population	**Metric**	**Derivation cohort April 2015- March 2017**		**Validation cohort April 2017- March 2019**	
		Area under ROC curve	Area under PR curve	Area under ROC curve	Area under PR curve
Total Respiratory	NEWS2	0.888 (0.881-0.895)	0.140	0.880 (0.873-0.887)	0.133
	Additive NEWS2 + Maximum previous 24hrs	0.902 (0.895-0.909)	0.144	0.898 (0.891-0.904)	0.144
Scale 1 Cohort	NEWS2	0.895 (0.887-0.903)	0.146	0.879 (0.871-0.887)	0.140
	Additive NEWS2 + Maximum previous 24hrs	0.907 (0.900-0.914)	0.150	0.896 (0.889-0.903)	0.150
Scale 2 Cohort	NEWS2	0.857 (0.838-0.877)	0.114	0.878 (0.859-0.897)	0.099
	Additive NEWS2 + Maximum previous 24hrs	0.880 (0.862-0.898)	0.118	0.903 (0.885-0.921)	0.122

In a balanced data set, where positive and negative outcomes are not dramatically skewed, it is useful to present data using ROC curves as they represent both positive and negative accuracy through use of sensitivity and specificity and thereby provide a more rounded view of performance. This has led to its widespread use in analysis of predictive scores. Area under the PR curve has not been extensively reported in studies assessing early warning scores, however it has been included here due to the negative skew of the outcome within this population, in common with all NHS hospital populations. In the respiratory population described in this chapter, death within 24 hours occurred in just

0.89% of vital signs observation sets. This could lead to an over-optimistic area under the ROC curve as ROC curves do not take into account incidence of outcome within the population, whereas it is included in PR curves through inclusion of positive predictive value.

Figure 4-6 Precision recall curve for NEWS2 score using outcome of death within 24 hours applied to total respiratory population

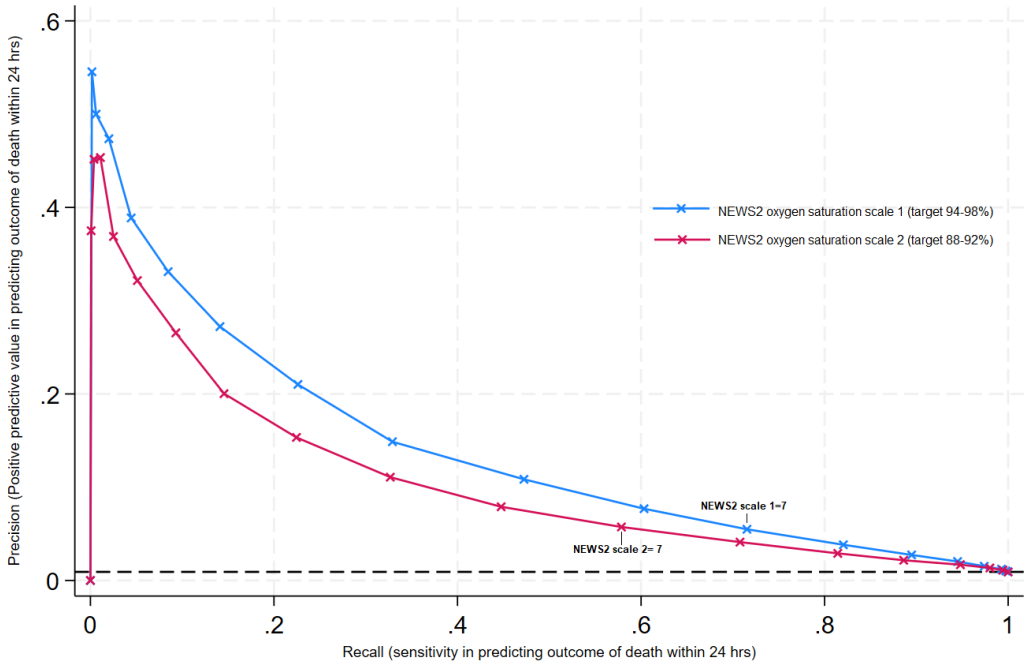


Baseline in PR curve graph is incidence of outcome within the population, in this case 0.009 (number of outcomes/number of observations)

In visual terms, precision recall curves have a different orientation to ROC curves, with a perfect curve being towards the top right of the graph. Figure 4-6 demonstrates the PR curve for NEWS2 with each cut point marked to demonstrate the values at specific cut points.

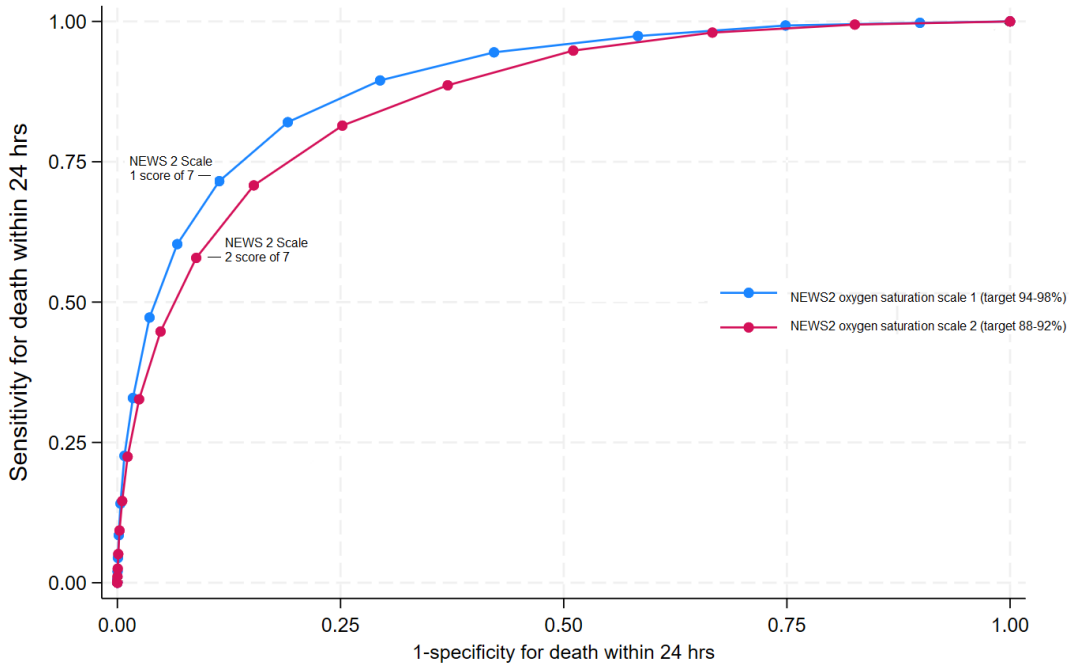
In order to demonstrate the impact of a small change in area under the PR curve in a population, Figure 4-7 uses patients without a diagnosis of COPD and plots the PR curves for both Scale 1 (area under the PR curve of 0.108) and Scale 2 (area under the PR curve of 0.114). These values for area under PRC curve suggest both scoring methods have poor discrimination for predicting death within 24 hours of an observation set, with scale 1 being superior to scale 2 in this population.

Figure 4-7 Precision recall curve for NEWS2 score applying saturation scale 1 versus saturation scale 2 using outcome of death within 24 hours applied to respiratory patients without a diagnosis of COPD



However, a different picture is projected using ROC curves to answer the same question. When the performance of NEWS2 is assessed in this way it suggests very good predictive ability for the outcome of death within 24 hours with values of 0.892 and 0.862 for scale 1 and scale 2 respectively as demonstrated in Figure 4-8. This is because the use of 1-Specificity in place of PPV removes incidence from the calculation. This demonstrates the importance of choice of analysis in evaluating a predictive score and understanding of both the population and clinical priorities.

Figure 4-8 ROC curves for NEWS2 applying scale 1 versus scale 2 using outcome of death within 24 hours applied to respiratory patients without a diagnosis of COPD



The values reported here are in line with other studies examining early warning scores. While it could be argued that PR curves would provide a more honest view of performance, because area under the ROC curve has been the predominant measure in the literature for two decades it was decided that this would be the most appropriate measure to use throughout the majority of the thesis. However it was felt that the limitations of this method should be made clear.

It is not possible to provide a direct match in cut points at every level of sensitivity and specificity between NEWS2 and the additive score described above as the ROC curves do not match in shape. However, the additive score demonstrates an improved performance at cut points equivalent to a NEWS2 of 5 and 7 in both the total respiratory population (Table 4-4) and in the scale2 cohort with a diagnosis of COPD (Table 4-5)

Table 4-4 Cut points for escalation with additive score combining maximum score in previous 24 hours and current NEWS2 matched to the NEWS2 score with equivalent sensitivity in the total respiratory population

2015-2017 Derivation Cohort						2017-2019 Validation Cohort					
Additive Score			NEWS 2			Additive Score			NEWS 2		
Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity
0	1.00	0.00	0	1.00	0.00	0	1.00	0.00	0	1.00	0.00
1	1.00	0.02				1	1.00	0.01			
2	1.00	0.06				2	1.00	0.04	1	1.00	0.09
3	1.00	0.12	1	1.00	0.12	3	1.00	0.09			
4	1.00	0.19				4	0.99	0.16			
5	0.99	0.27	2	0.99	0.27	5	0.99	0.23	2	0.99	0.24
6	0.99	0.35				6	0.99	0.31			
7	0.98	0.43	3	0.97	0.43	7	0.98	0.39			
8	0.96	0.51				8	0.97	0.47	3	0.97	0.40
9	0.95	0.58	4	0.93	0.60	9	0.95	0.55			

10	0.93	0.65				10	0.94	0.62	4	0.93	0.56
11	0.90	0.72				11	0.91	0.69			
12	**0.87**	**0.78**	**5**	**0.87**	**0.72**	**12**	**0.89**	**0.75**	**5**	**0.88**	**0.69**
13	0.83	0.83	6	0.80	0.82	13	0.85	0.80			
14	0.79	0.87				14	0.80	0.85	6	0.80	0.80
15	**0.72**	**0.90**	**7**	**0.68**	**0.90**	**15**	**0.74**	**0.88**	**7**	**0.70**	**0.88**
16	0.64	0.92				16	0.68	0.91			
17	0.57	0.94	8	0.57	0.94	17	0.61	0.94	8	0.59	0.93
18	0.50	0.96	9	0.45	0.97	18	0.53	0.95	9	0.46	0.96
19	0.43	0.97				19	0.44	0.97			
20	0.36	0.98	10	0.32	0.98	20	0.36	0.98	10	0.32	0.98
21	0.28	0.99				21	0.29	0.99			
22	0.22	0.99	11	0.20	0.99	22	0.23	0.99	11	0.23	0.99
23	0.16	1.00				23	0.18	0.99			
24	0.13	1.00	12	0.13	1.00	24	0.14	1.00	12	0.15	1.00

25	0.08	1.00	13	0.07	1.00	25	0.11	1.00			
26	0.06	1.00				26	0.09	1.00	13	0.09	1.00
27	0.04	1.00	14	0.04	1.00	27	0.06	1.00	14	0.04	1.00
28	0.02	1.00	15	0.02	1.00	28	0.04	1.00			
29	0.02	1.00				29	0.02	1.00	15	0.02	1.00
30	0.01	1.00				30	0.01	1.00			
31	0.01	1.00	16	0.00	1.00	31	0.01	1.00	16	0.00	1.00
32	0.00	1.00				32	0.00	1.00	17	0.00	1.00
33	0.00	1.00	17	0.00	1.00	33	0.00	1.00			

Table 4-5 Cut points for Additive NEWS2 score matched to NEWS2 score with closest matched sensitivity for death in 24 hours- Scale 2 cohort with a diagnosis of COPD

2015-2017 Derivation Cohort						2017-2019 Validation Cohort					
Additive Score			NEWS 2			Additive Score			NEWS 2		
Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity	Cut point	Sensitivity	Specificity
0	1.00	0.00	0	1.00	0.00	0	1.00	0.00	0	1.00	0.00
1	1.00	0.02				1	1.00	0.01	1	1.00	0.13
2	1.00	0.05				2	1.00	0.04			
3	1.00	0.12	1	1.00	0.14	3	0.98	0.11			
4	0.99	0.20				4	0.98	0.18			
5	0.99	0.28	2	0.97	0.30	5	0.98	0.27			
6	0.97	0.37				6	0.98	0.37			
7	0.95	0.46	3	0.95	0.47	7	0.97	0.46	2	0.97	0.30
8	0.93	0.55				8	0.95	0.56	3	0.95	0.47
9	0.91	0.63				9	0.93	0.65	4	0.90	0.66

10	0.87	0.71	4	0.87	0.64	10	0.89	0.73			
11	0.82	0.77				11	0.85	0.80			
12	**0.77**	**0.83**	**5**	**0.77**	**0.77**	**12**	**0.83**	**0.85**	**5**	**0.81**	**0.79**
13	0.70	0.87	6	0.68	0.87	13	0.76	0.89	6	0.71	0.88
14	0.66	0.91				14	0.70	0.92			
15	**0.58**	**0.94**	**7**	**0.53**	**0.93**	**15**	**0.63**	**0.95**	**7**	**0.57**	**0.94**
16	0.51	0.96				16	0.54	0.96			
17	0.44	0.97	8	0.43	0.96	17	0.42	0.98	8	0.42	0.97
18	0.38	0.98				18	0.35	0.98	9	0.28	0.99
19	0.32	0.99	9	0.32	0.98	19	0.25	0.99	10	0.19	0.99
20	0.28	0.99	10	0.24	0.99	20	0.17	0.99			
21	0.22	0.99				21	0.13	1.00	11	0.12	1.00
22	0.17	1.00	11	0.15	1.00	22	0.12	1.00			
23	0.11	1.00	12	0.10	1.00	23	0.10	1.00			
24	0.07	1.00	13	0.06	1.00	24	0.08	1.00	12	0.07	1.00

25	0.05	1.00				25	0.07	1.00			
26	0.03	1.00	14	0.03	1.00	26	0.05	1.00	13	0.05	1.00
27	0.02	1.00				27	0.03	1.00	14	0.03	1.00
28	0.01	1.00	15	0.01	1.00	28	0.02	1.00			
29	0.01	1.00				29	0.01	1.00	15	0.01	1.00
30	0.00	1.00	16	0.00	1.00	30	0.00	1.00			

4.3.6 Exploration of additional scoring component

It has been suggested that the addition of a graded FiO2 score as an additional category to future iterations of NEWS could improve risk prediction [102], creating a 'NEWS-FiO2' variant. In this population, application of a previously described NEWS-FiO2 did not provide significant improvement in area under the ROC curve in predicting outcome of death within 24 hours. However, this may be attributed to the small number of outcomes present in the study population. Both the original NEWS2 and NEWS-FiO2 demonstrated improvement in discrimination when maximum score in the preceding 24 hours was applied to the total respiratory population and Scale 2 cohorts (Table 4-6).

Table 4-6 Area under ROC and PRC for NEWS2, additive score combining NEWS2 and maximum score in the preceding 24 hours, NEWS-FiO2 and additive score combining current NEWS-FiO2 and maximum NEWS-FiO2 in the preceding 24 hours

Prediction of outcome of death within 24 hours of an observation		NEWS2	Additive NEWS2 + max NEWS2 in previous 24 hrs	NEWS2 FiO2	Additive NEWS2 FiO2 + max NEWS2- FiO2 preceding 24 hrs
All Respiratory 2015-2017	Area under ROC	0.888 (0.881-0.895)	0.902 (0.895-0.909)	0.890 (0.882-0.897)	0.901 (0.894-0.908)
	Area under PR curve	0.140	0.144	0.158	0.167
Scale 2 Cohort 2015-2017	Area under ROC	0.857 (0.838-0.877)	0.880 (0.862-0.898)	0.865 (0.847-0.884)	0.883 (0.866-0.900)
	Area under PR curve	0.115	0.118	0.123	0.132
All Respiratory 2017-2019	Area under ROC	0.880 (0.873-0.887)	0.898 (0.892-0.905)	0.887 (0.881-0.894)	0.900 (0.894-0.907)
	Area under PR curve	0.134	0.144	0.145	0.155
Scale 2 cohort 207-2019	Area under ROC	0.878 (0.860-0.897)	0.903 (0.885-0.921)	0.880 (0.861-0.899)	0.899 (0.881-0.918)
	Area under PR curve	0.100	0.121	0.102	0.128

4.4 Discussion

In our study, NEWS2 had good prognostic ability for predicting death within 24 hours in the overall respiratory population, but a reduced prognostic ability in patients with a diagnosis of COPD. We also created a simple additive model combining most recently recorded NEWS2 with maximum score in the preceding 24 hours that could be used to reduce the number of observations reaching the threshold for escalation without affecting sensitivity for predicting which observations would be followed by death within 24 hours. A similar improvement in prognostic accuracy was indicated if the same approach was applied to a score incorporating FiO2

Following the release of the original NEWS in 2012 there has been ongoing evaluation of the score with the result that a second oxygen scale and additions to the AVPU criteria were made for NEWS2. While Scale 2 mitigated concerns regarding hyperoxia in patients at risk of type 2 respiratory failure it did not account for other baseline characteristics of these patients which impact on the ability of the score to predict which patients are at risk of deterioration. In addition, patients admitted to hospital with COPD have a lower mortality than the overall respiratory population (4·0% vs 5·7% in the derivation cohort and 3·1% vs 5·5% in the validation cohort). This makes the positive predictive value even more important due to skew between observations and outcomes and thereby the potential for excessive workload and unnecessary intervention

Echevarria et al. [77] analysed the performance of NEWS2 scale 2 when applied to patients with COPD. Scale 2 led to a reduction in scores reaching escalation thresholds, improved discrimination when compared to the original NEWS score (area under ROC curve 0·72 vs 0·65) and did not fail to identify any outcomes escalated by scale 1. Pimentel et al. [54] used a combination of coding and oxygen prescriptions to identify patient cohorts at risk of hypercapnic respiratory failure and confirmed hypercapnic respiratory failure. The performance of NEWS and the scale 2 component of NEWS2 (the modified AVPU component of NEWS2 was not applied) was compared in these cohorts to respiratory patients without risk factors for hypercapnia. As in our study, NEWS2 had worse predictive ability in the cohort with hypercapnic respiratory failure. These findings, and ours, suggest that the underlying physiological changes from chronic respiratory disease make NEWS2 less effective in patients at risk, or with hypercapnic respiratory failure, including those with COPD.

Using trends in vital signs observations has been shown to improve predictive ability [55, 56]. In this study, novel variables created from the pattern of NEWS2 scores preceding the most recently recorded set of observations were demonstrated to be independent predictors of outcome, and

enhanced the prognostic ability of NEWS2 when combined with most recently recorded NEWS2 score in bivariate models.

This demonstrates the potential to further improve NEWS without having to change either the mode of data collection or the observations recorded, and providing additional value even where additional factors such as FiO2 are included. Furthermore, use of maximum score in the preceding 24 hours would be possible in a paper-based system, while additional modelling could potentially combine multiple variables to improve accuracy in an electronic system.

Our study is the first to examine the possible impact on demand as a surrogate of workload through utilisation of an additional layer of risk assessment. Applying a cut point of 12 to the additive model combining NEWS2 and maximum NEWS2 in the preceding 24 hours, corresponding in sensitivity to a NEWS2 score of 5, would result in 7035 (9·2%) fewer scores meeting the threshold for escalation in the overall population and 1366 (11·2%) fewer scores reaching the threshold for escalation in the scale 2 cohort with a diagnosis of COPD, without reducing sensitivity in predicting death within 24 hours (Table 4-6).

In conclusion, chronic pathophysiological changes, such as those found in respiratory disease, affect the prognostic ability of NEWS2. This prognostic ability can be improved without the need for additional changes in data collection or major changes to existing systems by addition of the maximum score in the preceding 24 hours to the most recently recorded NEWS2 and could be applied to future iterations of NEWS if other variables such as graded FiO2 were to be included; this approach could easily be tested in other centres. This simple and scalable improvement could have beneficial implications all healthcare systems which strive to balance the seesaw of resource limitations versus the need to predict, react to, and prevent clinical deterioration in hospitalized patients.

5 The National Early Warning Score in Context- A retrospective cohort study to analyse the impact of specialty and age on performance and the potential use of pattern to improve performance

5.1 Introduction

Having addressed the respiratory population, this third quantitative study was designed to further explore the one-size fits all approach of NEWS2. Hospital populations are heterogeneous with a wide range of age, pre-existing comorbidity and admission diagnoses [110]. Younger patients in particular have a lower in hospital mortality rate, can maintain normal physiological values despite pathology and are less likely to have chronic changes in physiology[111] that impact baseline vital signs observations and therefore baseline NEWS score. Given that the NEWS score is designed to identify perturbations from normal physiological ranges which may indicate clinical deterioration and an increased risk of deterioration or death, it is perhaps unsurprising that the performance of a uniform predictive score depends upon the population studied.

Previous studies have explored the impact of applying some patient factors, such as age, on the performance of the original NEWS and other scoring systems [112-114]. Applying adjustments for age to NEWS score on arrival in the emergency department setting has also been shown to improve prediction of in hospital mortality [112]. However, there has not been an analysis of the difference in statistical performance of NEWS2 across different age-groups and specialty areas. The primary objective of this study was to describe the impact of population characteristics such as age and diagnostic grouping on the performance of NEWS2. The secondary objective was to determine whether addition of a simple pattern variable could be used to improve performance [115].

5.2 Methods

5.2.1 Data source

The anonymised database of vital signs observations linked to demographic details and outcome described in chapter 2 was again used here. In this study, all adult admissions to Nottingham University Hospitals Trust between January 2016 and December 2019 were included.

5.2.2 Population

Admissions were initially separated into medical and surgical cohorts according to the specialty they were admitted under, to allow comparison. Additional categories were created using diagnostic grouping based on ICD10 coding at admission (see Table 5.1 incorporate into each category). This followed a similar approach to the NHS digital groupings used to present Summary Hospital Mortality Rate Indicator statistics by disease group. These groups were defined before analysis of performance was performed. The final groups were infection, neoplasm, blood, endocrine and metabolic, neurological and psychiatric, head and neck, cardiac, stroke and associated sequelae, vascular, respiratory and thoracic, gastrointestinal, skin, rheumatology, renal tract, genital and obstetric, congenital, miscellaneous, injuries and poisons, public health. Patients admitted under obstetrics were excluded as they use a different early warning score and have separate mechanisms for responding to deterioration. Patients coded with obstetric diagnoses but admitted under gynaecology and therefore managed using the protocols as general adult patients were retained in the genital and obstetric group per diagnosis, and the surgical cohort. Any observations coded as end of life care were removed from the analysis as the focus in these patients is palliation of symptoms rather than preventing deterioration.

Table 5-1 ICD 10 codes associated with each disease group[116]

Disease Group	**ICD 10 Codes**
Infection of parasite	All of A and B
Neoplasm	All of C and D0-D49
Blood Disorder	D50 onwards
Endo Metabolic	All of E
Neuro Psych	All of F And G
Head and Neck	All of H
Heart	I0-I52
Stroke	I60-I69
Vascular	I70-I99
Respiratory	All of J
Gastrointestinal system	All of K
Skin disorders	All of L
Rheumatology	All of M
Renal tract	N0-N39
Genital and Obstetric conditions (Early pregnancy conditions managed by gynaecology only)	N40-99 All of O that apply
Congenital	All of Q
Miscellaneous	All of R
Injuries, Poisons and Accidents	All of S, T, V, W, X and Y
Public Health	All of Z

5.2.3 Analysis

The primary outcome for all analyses was death within 24 hours of an observation set. Area under receiver operating curves for primary outcome were generated to allow comparison with previous research into the performance of NEWS2. Sensitivity and specificity at the clinically relevant cut point of 7 [75] was also calculated for each diagnostic and age grouping. The number needed to evaluate (NNE or workup detection ratio), i.e. the number of scores reaching threshold for escalation per outcome of death within 24 hours identified, was also calculated for each population and age grouping [97].

To assess the impact of age on ability of NEWS2 score to predict death within 24 hours, the population was initially divided into age-bands of 18-44, 45-64, 65-84 and 85 years or more based on age at admission. After initial analysis, this was streamlined to a single 65 year cut off. Further analysis was performed within the populations designated by specialty and diagnosis based on age under 65 or 65 years and over.

5.2.4 Incorporating pattern of NEWS2

In order to address the secondary objective, to ascertain whether any improvements to performance of NEWS2 could be made consistently across all population groups, the impact of incorporating different patterns of scoring on ability to predict outcome of death within 24 hours was tested. These included the maximum, minimum, standard deviation and mean score in the preceding 24 hours[117]. The patterns were used to create restricted cubic spline models with three knots, as indicated by the data and to reduce the risk of over fit, at the placement recommended as by Harrell [109]. Univariate models were created using the uvrs package in STATA. Each variable was then combined with current NEWS2 score using the mvrs package to create bivariate restricted cubic spline models. The best performing spline models were fit as additive models to allow use in less sophisticated systems, as some NHS trusts still use paper charting. Predictive additive models were created using maximum NEWS2 score in the preceding 24 hours combined with most recent NEWS2 score and mean NEWS2 score in the preceding 24 hours combined with most recent NEWS2 score as these were the best performing of the spline models.

Cut points with equivalent sensitivity for predicting primary outcome were applied to the best performing score to allow prediction of NNE for comparison with NEWS2.

Results

Between 1st January 2016 and 31st December 2019, there were 354,830 adult inpatient admissions. Median age at admission was 61 years. 198,300 episodes (56%) were admitted under medical specialties and 125,604 (35%) under surgical specialties. The breakdown of admission episodes, deaths and observations within each of the disease groups is detailed in table 5-2 below. The in-hospital mortality rate within these disease groups varied significantly from 0 to 10.9%.

Table 5-2 Number and proportion of deaths and observations followed by death within 24 hours by disease group.

	Median age	Died in hospital	% died in hospital	Observations not followed by death in 24 hours	Observations followed by death within 24 hours	% of observations followed by death within 24 hours
Stroke	76 (51-81)	714	10.5	411,309	1,583	0.38
Infection or parasite	66 (46-81)	1,627	9.8	681,126	5,765	0.84
Respiratory and Thoracic	71 (56-81)	2,667	7.8	1,029,705	10,880	1.05
Neoplasm	66 (51-71)	1,549	5.2	953,891	3,197	0.33
Heart	71 (56-81)	969	3.6	646,859	4,521	0.69
Vascular	71 (56-81)	205	3.2	148,650	600	0.40
Gastrointestinal system	56 (41-71)	899	2.9	683,834	3,047	0.44

Renal tract	66 (46-76)	411	2.5	363,533	1,292	0.35
Endo Metabolic	61 (36-76)	179	2.3	198,154	511	0.26
Miscellaneous	61 (41-76)	968	1.7	942,380	2,996	0.32
Injuries Poisons Accidents	61 (36-76)	690	1.7	1,089,407	2,701	0.25
Neuro Psych	51 (36-71)	196	1.2	331,227	508	0.15
Blood disorder	56 (41-76)	59	1	88,168	250	0.28
Skin Disorders	51 (31-71)	78	0.9	154,589	321	0.21
Rheumatology	61 (46-76)	161	0.6	561,314	550	0.10
Public Health	61 (51-71)	9	0.2	43,745	17	0.04
Genital and Early Obstetrics	31 (21-51)	7	0.0	122,030	28	0.02

5.2.5 Performance of NEWS2 in different populations

NEWS2 performance varied by primary diagnoses, as shown in Table 5-3. Sensitivity in predicting outcome of death within 24 hours of an observation set at a NEWS threshold score of 7 and above ranged from 35.7% to 68.4%. The NNE was not observed to be linked to area under the ROC curve, that is to say higher area under ROC curve was not always associated with a lower NNE. For example in patients with a rheumatology diagnosis the NNE at a threshold score of 7 or more was 21.3 (95% CI 19.4-23.6) despite an area under the ROC curve of 0.934. This compared to patients with a diagnosis relating to an infection or parasite with a lower area under ROC curve of 0.909 but also a lower NNE at 13.2 (95% CI 12.8-13.7). This is due to the relationship between sensitivity, specificity and prevalence. As NNE is based on the PPV of a score, as opposed to area under the ROC curve which does not take into account prevalence, a low prevalence would impact NNE but not area under ROC curve.

Table 5-3 Area under the ROC curve, sensitivity, specificity and number needed to evaluate for NEWS2 in predicting outcome of death in 24 hours (disease groups displayed limited to 10 with highest mortality- full table included as appendix table 5-1)

	AUROC	Sensitivity at 7	Specificity at 7	NNE at 7 (95% CI)
Total population	0.916	60.4	96.5	14.9 (14.7-15.1)
Medical	0.909	61.2	95.6	14.2 (14.0-14.3)
Surgical	0.914	55.4	98.2	20.2 (19.5-20.9)
Stroke	0.896	58.0	96.2	18.9 (17.8-20.1)
Infection or parasite	0.909	65.5	93.7	13.2 (12.8-13.7)
Respiratory and Thoracic	0.890	68.4	90.5	14.5 (14.2-14.8)
Neoplasm	0.921	57.8	97.6	14.4 (13.8-15.0)
Heart	0.879	53.4	95.6	13.6 (13.1-14.2)
Vascular	0.850	36.3	98.0	15.9

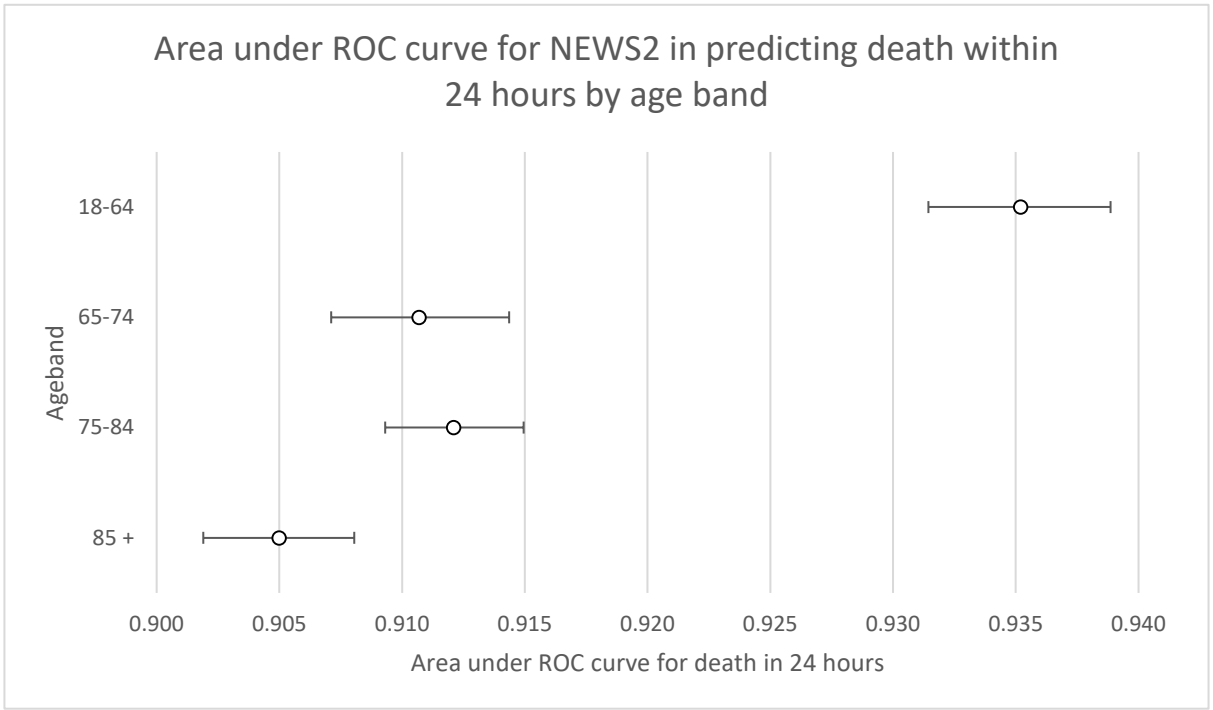
				(14.0-18.0)
Gastrointestinal system	0.907	56.6	97.5	12.6 (12.0-13.1)
Renal tract	0.897	47.7	98.0	14.6 (13.5-15.6)
Endo Metabolic	0.880	51.7	97.1	24.6 (21.9-27.7)
Miscellaneous	0.919	54.7	97.8	14.7 (14.1-15.4)

NNE= Number Needed to Evaluate, i.e. the number of scores reaching threshold for every score followed by an outcome within 24 hours

5.2.6 Impact of age on NEWS2

Area under the ROC curves for predicting death within 24 hours by age-bands aligned to those used by the National Cardiac Arrest Audit demonstrated a visible difference in performance of NEWS2 between the 18-64 age band and the 3 bands above age 65 (Figure 5-1).

Figure 5-1 Area under the ROC curve for NEWS2 in the total population when split by age at admission into NCAA age bands (error bars represent 95% CI)



Using a cut point of 65 demonstrates a considerable difference in mortality in the overall population and in the different subpopulations defined by specialty and disease group, as shown in Table 5-4 below.

Table 5-4 Mortality rates defined by age with cut point of 65 years (disease groups displayed limited to 10 with highest mortality- full table included as appendix table 5-2)

	Mortality n (%) in patients aged under 65	Mortality n (%) in patients aged 65 and older
Total population	2,399 (1.0)	12,253 (5.7)
Surgical patients	266 (0.3)	1,429 (2.1)
Medical Patients	2,133 (1.6)	10,808 (7.6)
Stroke	99 (3.4)	793 (13.8)
Infection or parasite	292 (3.2)	1,637 (15.0)
Respiratory and Thoracic	347 (2.1)	3,087 (11.7)
Neoplasm	673 (3.5)	1,301 (6.8)
Heart	135 (1.1)	1,120 (5.1)
Vascular	38 (1.3)	216 (4.0)
Gastrointestinal system	291 (1.2)	870 (5.5)
Renal tract	55 (0.5)	491 (4.7)
Endo Metabolic	52 (0.9)	180 (4.2)
Miscellaneous	208 (0.5)	1,020 (3.0)

5.2.7 Applying pattern to improve performance of NEWS2

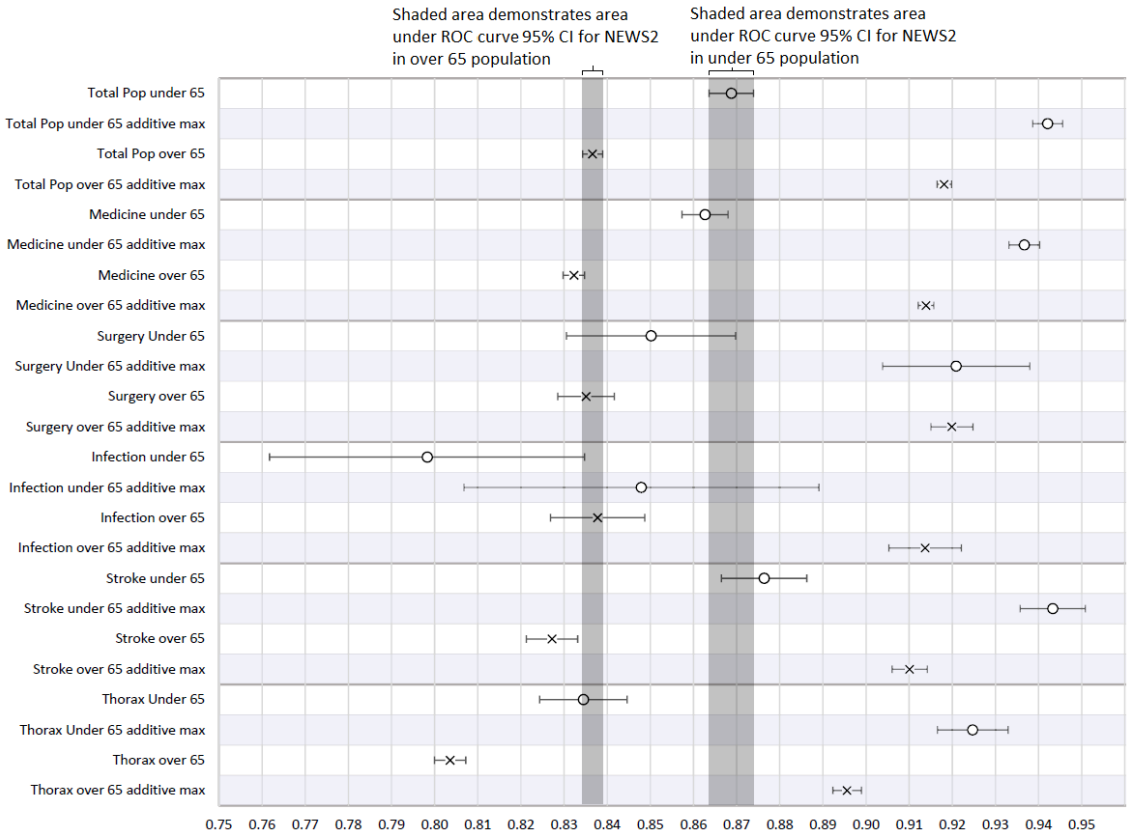
There was minimal difference in performance between scores created using restricted cubic splines and those created using a simple additive method (Table 5-5). The two best performing scores were those combining most recent NEWS2 with mean score in the preceding 24 hours and maximum score in the preceding 24 hours[118]. As one of the key aims of this study is to identify potential improvements to NEWS2 that could be used in all hospitals without updating infrastructure, the additive score combining most recent NEWS2 and maximum score in the preceding 24 hours was taken forward.

Table 5-5 Area under ROC curve for NEWS2 and scores combining current NEWS2 with patterns of score in the preceding 24 hours including maximum, minimum, standard deviation and mean. The first four models were created using restricted cubic splines, the final two using an additive approach

	NEWS2	Spline current NEWS2 + Maximum value preceding 24 hours	Spline current NEWS2 + Mean Value preceding 24 hours	Spline current NEWS2 + SD of values preceding 24 hours	Spline current NEWS2 + Minimum value preceding 24 hours	Additive score of current NEWS2 + maximum value preceding 24 hours	Additive score of current NEWS2 + mean value preceding 24 hours
	AUROC	AUROC	AUROC	AUROC	AUROC	AUROC	AUROC
Total	0.916	0.926	0.926	0.921	0.920	0.925	0.926
Medicine	0.909	0.918	0.918	0.913	0.913	0.920	0.920
Surgery	0.914	0.925	0.925	0.920	0.919	0.924	0.924
Infection or	0.909	0.916	0.917	0.912	0.913	0.918	0.921
Neoplasm	0.921	0.929	0.929	0.926	0.926	0.928	0.931
Blood disorder	0.892	0.899	0.896	0.895	0.892	0.916	0.913
Endo Metabolic	0.880	0.895	0.892	0.885	0.884	0.903	0.902
Neuro Psych	0.923	0.929	0.930	0.925	0.925	0.935	0.936
Heart	0.879	0.893	0.891	0.887	0.884	0.893	0.891
Stroke	0.896	0.904	0.905	0.898	0.903	0.908	0.910
Vascular	0.850	0.865	0.864	0.863	0.858	0.859	0.859
Respiratory and Thoracic	0.890	0.901	0.899	0.894	0.894	0.901	0.900
Gatrointestinal system	0.907	0.920	0.918	0.914	0.911	0.925	0.926
Skin Disorders	0.928	0.937	0.937	0.931	0.931	0.938	0.938
Rheumatology	0.934	0.946	0.944	0.942	0.938	0.957	0.959
Renal tract	0.897	0.911	0.910	0.901	0.901	0.902	0.903
Genital and Obs	0.981	0.987	0.990	0.983	0.987	0.985	0.990
Miscellaneous	0.919	0.926	0.926	0.923	0.922	0.922	0.924
Injuries Poisons Accidents	0.923	0.934	0.934	0.927	0.927	0.936	0.936
Public Health	0.886	0.884	0.905	0.901	0.912	0.857	0.892

Figure 5-2 demonstrates that the novel additive score combining most recent NEWS 2 and maximum score in the preceding 24 hours produced a larger area under the ROC curve than the original NEWS2 for predicting the outcome of death in 24 hours in all diagnostic groups. This was also the case when split into patients aged under 65 and those aged 65 and over.

*Figure 5-2 area under the ROC curve and number needed to evaluate for NEWS2 versus the combination of current NEWS2 and maximum NEWS2 in preceding 24 hours**



* Each population group is split into those aged over 65 and under 65, with NEWS2 unshaded and the additive maximum score in grey. Error bars represent 95% confidence intervals. The baseline is represented by the NEWS2 score in the total population in those aged a) over 65 and b) those aged under 65 with 95% confidence intervals for these groups represented by the vertical shaded areas. For additional diagnostic groups see supplemental figure 1.

Area under the ROC curve does not represent the full story as cut points are required to trigger actions. Table 5-6 shows the sensitivity, specificity and NNE in the score combining maximum score in the preceding 24 hours and most recent NEWS2 score. The cut point of 15 represents the cut point at which sensitivity for death within 24 hours is most closely matched to a NEWS2 cut point of

7. This demonstrates that in most groups the sensitivity for outcome using the additive score can improve on the sensitivity for NEWS without reducing specificity or increasing the NNE.

Table 5-6 A comparison of sensitivity, specificity and NNE for death within 24 hours at a NEWS cut point of 7 and an additive score cut point of 15 with disease groups ordered by mortality rate.

	Sensitivity NEWS2 7	Specificity NEWS2 7	NNE NEWS2 7 (95% CI)	Sensitivity Additive score 15	Specificity Additive score 15	NNE Additive score 15
Total population Under 65	0.63	0.97	26.6 (25.8-27.5)	0.65	0.97	26.0 (25.2-26.8)
Total population Over 65	0.60	0.96	12.7 (12.5-12.8)	0.62	0.96	12.6 (12.4-12.7)
Medical population Under 65	0.64	0.96	23.4 (22.6-24.1)	0.66	0.96	23.2 (22.5-24.0)
Medical population Over 65	0.61	0.95	12.3 (12.2-12.5)	0.62	0.96	12.3 (12.2-12.5)
Surgical population Under 65	0.59	0.99	62.8 (56.4-69.9)	0.60	0.99	58.7 (52.6-65.4)
Surgical population Over 65	0.55	0.98	14.9 (14.4-15.5)	0.58	0.98	14.1 (13.7-14.7)
Stroke under 65	0.57	0.97	44.9 (37.0-54.5)	0.53	0.96	54.3 (44.1-64.0)
Stroke over 65	0.58	0.96	16.1 (15.1-17.1)	0.59	0.96	17.1 (16.0-18.2)
Infection under 65	0.72	0.95	18.9 (17.6-20.2)	0.76	0.94	19.4 (18.2-20.8)
Infection over 65	0.64	0.93	11.7 (11.3-12.1)	0.66	0.93	12.3 (11.8-12.7)
Thorax Under 65	0.71	0.92	31.9 (29.8-34.1)	0.75	0.92	30.6 (28.6-32.6)
Thorax over 65	0.68	0.90	12.5 (12.3-12.8)	0.71	0.90	12.3 (12.1-12.6)
Neoplasm under 65	0.60	0.98	18.6 (17.3-20.0)	0.62	0.98	19.0 (17.7-20.5)
Neoplasm over 65	0.56	0.97	12.1 (11.5-12.7)	0.58	0.97	11.8 (11.2-12.5)
Heart under 65	0.57	0.96	24.2 (21.8-26.9)	0.63	0.96	22.9 (20.7-25.3)
Heart over 65	0.53	0.95	12.1 (11.6-12.5)	0.55	0.95	11.7 (11.3-12.2)
Vascular Under 65	0.41	0.98	22.1 (16.8-29.2)	0.39	0.98	22.5 (17.0-30.0)
Vascular over 65	0.35	0.98	14.3 (12.4-16.4)	0.35	0.98	14.5 (12.6-16.7)
Gastrointestinal under 65	0.60	0.98	23.5 (21.3-25.9)	0.61	0.98	23.4 (21.2-25.8)
Gastrointestinal over 65	0.56	0.97	9.6 (9.1-10.0)	0.55	0.97	9.8 (9.4-10.3)
Renal Under 65	0.43	0.99	40.7 (32.1-51.8)	0.47	0.98	39.6 (31.3-50.2)

Renal over 65	0.48	0.98	12.0 (11.1-12.9)	0.47	0.98	12.2 (11.4-13.2)
Endo Metabolic under 65	0.53	0.97	57.6 (44.4-74.9)	0.59	0.97	48.6 (38.2-62.0)
Endo Metabolic over 65	0.51	0.97	16.1 (14.1-18.3)	0.52	0.97	16.2 (14.2-18.5)
Miscellaneous under 65	0.56	0.98	21.2 (19.1-23.5)	0.53	0.98	20.7 (18.6-23.0)
Miscellaneous over 65	0.55	0.98	13.2 (12.6-13.9)	0.54	0.98	13.0 (12.3-13.7)

5.3 Discussion

In this study, NEWS2 was shown to have substantial variation in ability to predict death within 24 hours when applied across different inpatient populations defined according to specialty cohort, medical versus surgical inpatients, age grouping, eventually applying 65 as a cut point, and diagnostic grouping according to admission ICD10. This is important because NEWS2 is applied uniformly across all of the groups studied, with the exception of the oxygen saturation scale 2 for use in patients with type 2 respiratory failure.

Mortality rates varied greatly in the subpopulations examined. Within the study population as a whole, when examined using an age cut point in patients aged under 65 the mortality was 0.99% and in those aged over 65 it was 5.65%. When the subpopulations were defined by admission diagnosis mortality varied between 0-10.5%. As the prevalence of outcome is associated with the level of workload generated per outcome identified it is important to consider this difference in mortality in combination with other metrics. Clinicians have previously suggested that in patients with older age, NEWS2 is not able to account for the impacts of comorbidity, polypharmacy and conditions which are not immediately evidenced by vital sign derangement [119]. In our analyses of area under ROC curve, sensitivity and specificity at clinically relevant cut of 7, which is the threshold for escalation to the registrar, we demonstrated that both age and disease group were associated with substantial variation in performance of NEWS2 in predicting outcome of death within 24 hours.

Area under the ROC curve, sensitivity and specificity were all higher in patients under 65 irrespective of diagnosis. This suggests different cut points could be beneficial for those aged over and under 65. This is particularly pertinent in patients aged over 65 as this group combines lower sensitivity with higher mortality rate. This is consistent with previous studies demonstrating age to be an important factor in the performance of early warning scores based on vital signs alone. An earlier study looking at the relationship between deranged physiology and risk of mortality in different age groups demonstrated that deranged vital signs were more likely to be associated with mortality in older patients. In patients aged 80 years or older (n=3201) with a respiratory rate of 24-25 breaths a minute there was a four-fold increase in mortality compared to those aged 40-64 years (n=2585) with the same respiratory rate. The same was true of blood pressure, with those aged 80 or over having 10 times the mortality of those aged 40-64[113]. This relationship between increasing mortality with age, in the setting of a similar level of physiological derangement, was translated to an early warning score in a study examining mortality by age group in a cohort of Danish emergency admissions. This multi-centre study (n=19,123) demonstrated higher mortality in those aged over 60

in all early warning score categories when compared to younger patients[120]. The ability of admission NEWS to predict in hospital mortality in a further Danish multicentre cohort (n=14,809) and the Netherlands Emergency department Evaluation Database (n=50,448) found that the area under the ROC curve in the Danish cohort of the study reduced from 0.82 in those aged under 65 to 0.78 in those aged 65-80 and those aged over 80. In the Netherlands cohort area under the ROC curve for predicting in-hospital mortality from admission NEWS2 was 0.80 (95%CI 0.76-0.80 in those aged under 65, 0.75 in those aged 65-80 and 0.72 in those aged over 80 [112]. Despite the NEWS2 being more strongly associated with mortality in younger patients, due to the substantially lower prevalence of mortality in this group, the number needed to escalate is actually higher than in those patients aged over 65 at the cut point of 7 used as a threshold for escalating to the registrar in most centres. The link between NNE and prevalence of mortality is also the reason that NNE was lower in medical than surgical patients.

The variation by age in area under the ROC curve is also likely to be related to greater physiological reserve in younger patients. This is indicated by previous research showing increased risk of mortality with increasing age following MET calls [121]. However, this could only be determined through accurate documentation of every intervention as well as outcome. This was not possible using the data available for this study but is potentially feasible in the setting of electronic prescribing where interventions such as fluid and antibiotics would be timestamped and could be linked to observation data.

Variation in performance of early warning scores between different patient specialty cohorts, as demonstrated here, has been previously noted [51, 118]. However this is the first time we are aware of the performance of NEWS2 being examined across different diagnostic groups in this way. The observed variation is of particular importance in the respiratory population where reduced ability to predict death in 24 hours is combined with a population with relatively high mortality in comparison to the hospital inpatient population as a whole [54].

Addition of a pattern variable improved the prognostic ability in all groups compared to NEWS2 alone. This allowed an increase in sensitivity for the outcome of death within 24 hours to be achieved in most groups without seeing a decrease in specificity or major change in the NNE. This suggests that maximum score could be used to improve risk detection without having significant workload implications.

The strength of the study is the completeness and granularity of the data, the ability to assign diagnostic groups based on ICD 10, and by the large number of outcomes facilitating a more granular analysis without loss of power. The presence of multiple years also reduced the likelihood of data

being influenced by a single abnormal year of data. However, the data is from the single centre nature of the data and the inability to see the impact of any interventions triggered by a high score.

5.4 Conclusion

Age and diagnosis group were both demonstrated to be associated with mortality within 24 hours in this population. When combined with data from previous studies, this suggests that age and diagnostic differences should be considered when developing early warning scores. The improvement in ability to predict death within 24 hours when applying a pattern variable in addition to most recent NEWS suggests this approach could be used to improve detection of patients at risk of deterioration. Addition of maximum score in the preceding 24 hours is a straightforward metric that could be used in all systems, whether computer or paper based.

5.5 Appendix

Appendix Table 5- 1 Area under the ROC curve, sensitivity, specificity and number needed to evaluate for NEWS2 in predicting outcome of death in 24 hours (All disease groups)

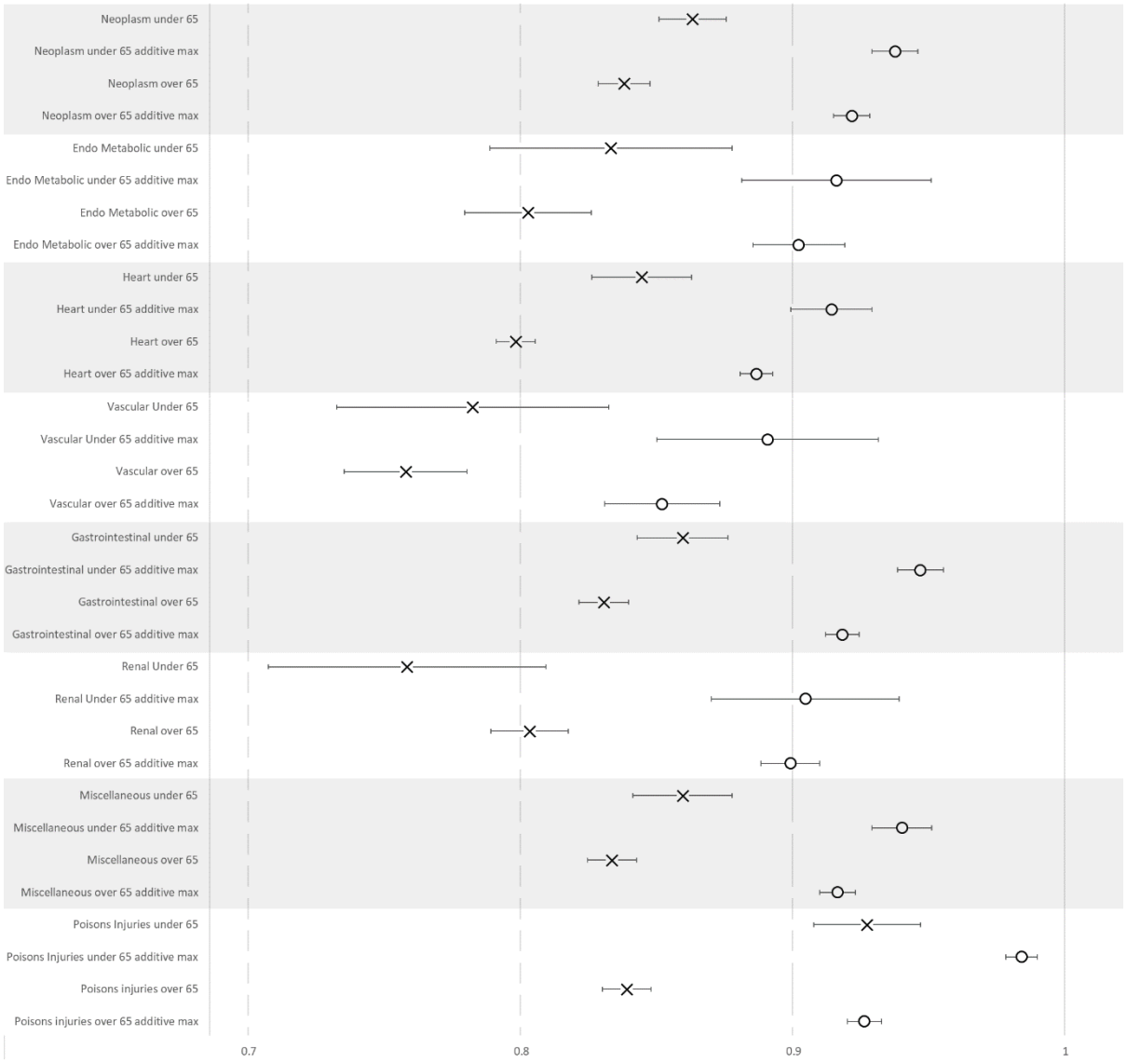
	AUROC	Sensitivity at 7	Specificity at 7	NNE at 7
Total population	0.916	60.4	96.5	14.8 (14.7-15.1)
Medical	0.909	61.2	95.6	14.2 (14.0-14.3)
Surgical	0.914	55.4	98.2	20.1 (19.5-20.9)
Stroke	0.896	58.0	96.2	18.9 (17.8-20.1
Infection	0.909	65.5	93.7	13.2 (12.8-13.7)
Respiratory and Thoracic	0.890	68.4	90.5	14.5 (14.2-14.8)
Neoplasm	0.921	57.8	97.6	14.4 (13.8-15.0)
Heart	0.879	53.4	95.6	13.6 (13.1-14.2)
Vascular	0.850	36.3	98.0	15.9 (14.0-18.0)
Gastrointestinal system	0.907	56.6	97.5	12.6 (12.02-13.1)
Renal tract	0.897	47.7	98.0	14.6 (13.6-15.6)
Endo Metabolic	0.880	51.7	97.1	24.6 (21.9-27.7)
Miscellaneous	0.919	54.7	97.8	14.7 (14.1-15.4)

Injuries Poisons Accidents	0.923	59.9	97.9	16.6 (15.8-17.4)
Blood disorder	0.892	42.4	98.1	19.5 (16.2-23.3)
Neuro Psych	0.923	59.4	98.0	23.6 (21.2-26.2)
Skin Disorders	0.928	53.0	98.0	18.4 (16.9-21.0)
Rheumatology	0.934	63.7	98.8	21.3 (19.4-23.6)
Public Health	0.886	35.7	99.3	78 (33.0-187.5)
Genital and Obs	0.981	39.3	99.3	92.4 (51.4-166.5)

Appendix Table 5- 2 Mortality rates defined by age with cut point of 65 years (all disease groups)

	Mortality n (%) in patients aged under 65	Mortality n (%) in patients aged 65 and older
Total population	2,399 (1.0)	12,253 (5.7)
Medical	266 (0.3)	1,429 (2.1)
Surgical	2,133 (1.6)	10,808 (7.6)
Stroke	99 (3.4)	793 (13.8)
Infection	292 (3.2)	1,637 (15.0)
Respiratory and Thoracic	347 (2.1)	3,087 (11.7)
Neoplasm	673 (3.5)	1,301 (6.8)
Heart	135 (1.1)	1,120 (5.1)
Vascular	38 (1.3)	216 (4.0)
Gastrointestinal system	291 (1.2)	870 (5.5)
Renal tract	55 (0.5)	491 (4.7)
Endo Metabolic	52 (0.9)	180 (4.2)
Miscellaneous	208 (0.5)	1,020 (3.0)
Injuries Poisons Accidents	104 (0.4)	960 (4.1)
Blood disorder	15 (0.3)	63 (1.9)
Neuro Psych	47 (0.3)	215 (3.0)
Skin Disorders	10 (0.1)	101 (3.0)
Rheumatology	30 (0.2)	175 (1.1)
Public Health	1 (0.0)	10 (0.3)
Genital and Obs	1 (0.0)	7 (0.3)

Appendix figure 5. 1 Area under ROC of NEWS and additive Max score for predicting death within 24 hours- remaining diagnoses not included in figure 5.2 within the main text of the chapter



6 Nurse Concern- A Critical Decision Methods Study Of Nurse Concern In The Setting Of Concurrent Early Warning Score Use.

Having focussed on high volume data analysis in the first three studies included in this thesis, this chapter takes a qualitative approach to explore other factors applied in recognition of patient deterioration.

6.1 Introduction

The role of nursing staff in evaluating patients and highlighting concerns to the medical team has long been the subject of investigation[122]. Following the widespread introduction of early warning scores from the late 1990s onwards, there has been a divergence in practice between systems which include nurse concern and other external data such as blood tests in their protocols, such as the Australian 'Between the Flags' model[123], and those which escalate based on vital signs alone, such as the National Early Warning Score (NEWS2)[94]. The data presented in chapter 2 demonstrated the impact of introducing a universal risk stratification system based entirely on vital signs, i.e. NEWS2, without any mitigation for degree of end of the bed clinical concern or patient specific factors such as age and comorbidity. The importance of such mitigation is made apparent in chapters 3 and 4 where a wide variation in the statistical accuracy of NEWS2 in predicting risk of death within 24 hours is noted when considering differences in age and primary diagnosis.

In the UK, where NEWS2 is now mandated for use in screening for risk of deterioration in hospital, research has largely focussed on the statistical accuracy of the score, the degree to which nursing staff adhere to protocols and what factors lead to deviation from those protocols[29, 124, 125]. However, several studies have suggested that vital sign deterioration is a late sign suggestive as it is of organ dysfunction, and that nursing assessment of a patient can detect deterioration up to 24 hours before their vital signs become sufficiently deranged to trigger a review[126]. In the data set used for chapters 2-4, it was discovered that between 7-17% of escalations to the registrar each year were made in patients not meeting the escalation criteria for the early warning score in place at the time. This breaks down as 6-16% in medical patients and 9-22% in surgical patients. Further work has suggested that despite the perception of nursing intuition as a nebulous concept there may be common factors which are analogous to, but not captured by, the vital signs measured by NEWS2[127].

However, nursing concern cannot be taken in isolation from the setting in which it operates. The ability for a nurse to evaluate a patient and the degree to which that opinion is utilised in making patient care decisions relies on a complex framework of interactions which need to be understood before changes can be considered to the current NEWS2 protocols.

We therefore set out to determine the factors involved in nurse assessment of a patient , what patient factors they use intentionally and automatically in the course of their patient care to assess condition at the end of the bed, and the perceived enabling and opposing factors in their ability to communicate their assessment to the appropriate people to act.

6.2 Methods

6.2.1 Design

A study was conducted in Nottingham University Hospitals Trust, a large acute NHS trust comprising 2 hospitals with tertiary inpatient care over a 10 month period from September 2021 to July 2022. Qualitative data were generated in the form of semi-structured interviews employing a combination of critical decision method [128, 129] approach and exploratory discussion around theme of bedside evaluation of patient and process of escalation for review and management. The critical decision method was chosen as it has been recognised as an effective approach to elicit expert knowledge. It allows a specific incident to be explored with the decision making process teased out and mapped in a manner that avoids leading the subject. The semi-structured nature of the questioning means no pre-existing assumptions are needed and all aspects of the participants experience can be explored. Other methods were considered, however a more structured approach would limit the scope for novel findings, while it has also been demonstrated that simulated case studies lack the complexity to describe a clinical situation and the low frequency of the events being studied make observation inefficient.

The participants were therefore asked to think of situations in their experience when they had either been worried about a patient with a relatively normal set of vital signs, or low early warning score where one was employed; and conversely where they were not worried about a patient with a relatively high early warning score, indicating more abnormal vital signs. The reasons for their clinical assessments in each of these circumstances was then explored, with questions to probe the reasoning behind their decision-making. Having considered these two examples, a more general question was asked about whether there were any additional factors, not present in the cases discussed, that would make them more concerned about a patient that are not currently captured by NEWS2. With the follow up of any bedside or patient factors that would reassure them.

The data once collected were then analysed using an emergent themes approach. Outputs from this analysis were then fed into a Systems Engineering Initiative for Patient Safety (SEIPS) system to illustrate the processes involved in the system of recognising and responding to patient deterioration across the hospital. This is a human factors model which illustrates how patients, staff, tasks, tools, the physical environment and the overall organisation in which they exist interact to influence processes and outcomes[130]. The work systems included in SEIPS can be described using a 'PETT' model, which examines the people, environment, tasks and tools involved[131] and their interactions in order to identify the interaction between different elements in identifying and escalating patients.

6.2.2 Participants

Participants were recruited from across all specialties with support from senior nursing staff to ensure engagement through email, physical literature in visible positions on the wards and engagement at nursing handover to ensure understanding of the aims of the project. The onus was then on potential participants to contact the study team at which point participant information sheets and consent forms were emailed to those indicating interest.

6.2.3 Ethical considerations

The project was approved by the University of Nottingham Medical Research Ethics Committee and NUHT Information governance and Research and Innovation teams. Consent forms were signed digitally prior to the interview with confirmation of understanding and ongoing consent at the start of each interview. Consent forms emphasised that participants could withdraw at any time, that all participation would be kept anonymous and that recordings would only be seen by the study team. As per the information governance team, all study documentation which it was not possible to anonymise, including recordings and consent forms, were kept in a secure folder on the trust server.

6.2.4 Data Analysis

Transcribed interviews were stored and analysed in NViVO12. Coding was completed in 2 phases. With initial open coding to establish patterns before organising emergent themes into more structured categories.

6.3 Results

A total of 17 interviews were carried out. Of these, 7 participants had worked exclusively in medical specialties, 8 had experience of both medical and surgical specialties, and 2 had worked exclusively in surgical specialties (see Table 6-1). Interviews lasted between 25 and 80 minutes.

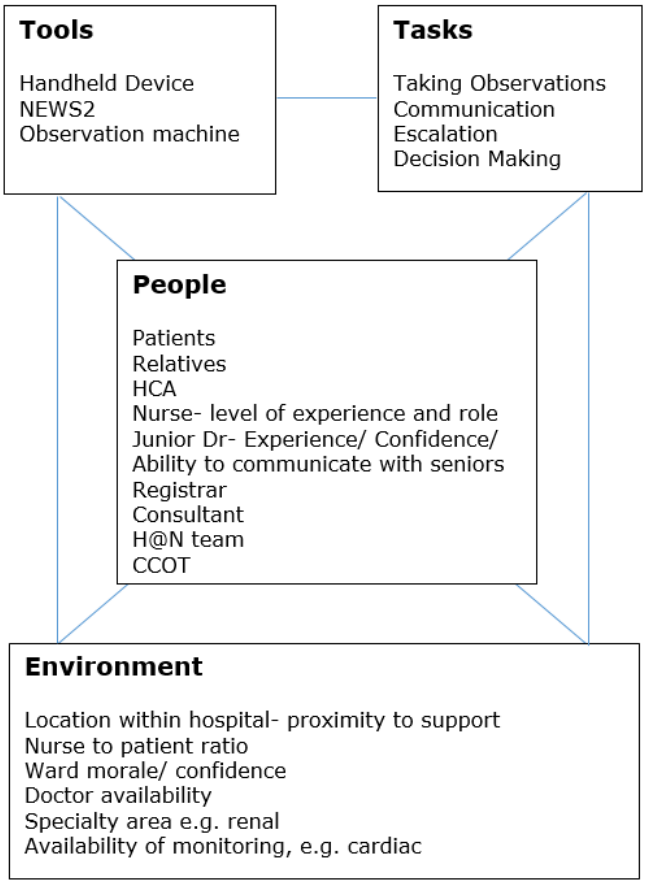
Table 6-1 Participants and their clinical backgrounds

	Years Qualified	Overriding specialty group	Gender	1st Specialty	2nd Specialty	3rd Specialty	4th Specialty
001	10+	Mixed	Female	General Medicine	General Surgery	Intensive Care	
002	10+	Medicine	Female	HCOP	Acute Medicine	Cardiology	
003	5-10	Mixed	Female	T&O	Theatre Recovery	Oncology	
004	1-4	Mixed	Female	Gynaecology	GUM		
005	10+	Medicine	Female	Renal	Not Applicable		
006	5-10	Mixed	Female	Gynaecology	Oncology	Breast	
007	5-10	Medicine	Male	Renal	Oncology		
008	10+	Medicine	Female	Haematology			
009	10+	Medicine	Female	Renal	General Medicine	CCOT	
010	10+	Medicine	Female	Cardiology			
011	10+	Medicine	Female	Oncology	Cardiology		
012	10+	Mixed	Female	T&O	General Surgery	Respiratory	Midwifery
013	5-10	Surgery	Female	Upper GI/ Colorectal Surgery	T&O	Pain	
014	10+	Surgery	Female	Neurosurgery			
015	10+	Mixed	Female	Neurosurgery	Oncology		
016	10+	Mixed	Female	T&O	Intensive Care		
017	10+	Mixed	Female	T&O	HCOP	Cardiology	Respiratory

* HCOP= Health Care of the Older Patient; T&O= Trauma and Orthopaedics; GUM= Genitourinary Medicine; GI= Gastrointestinal

Multiple themes were identified and structured to describe the interactions that were perceived to influence patient management using the SEIPS PETT model displayed in Figure 6-1.

Figure 6-1 PETT model displaying the people, environment, tasks and tools that contribute to patient care.



6.3.1 The factors which were found to influence recognition of patient deterioration are described below according to their category in the SEIPS model.

6.3.1.1 People

People form the hub of the SEIPS model. Within this category, the interviews identified two broader groups whose characteristics influence the assessment of a patient's clinical status, patients and staff. Within staff, the interaction between healthcare assistants, nursing staff and doctors was highlighted as being key to both identification and management of deterioration.

Patient Factors

Currently the only patient data used by NEWS 2 to predict risk of deterioration are vital signs collected at the bedside. However, these do not capture any of the following cues described by participants in their bedside assessment of patient stability or lack thereof.

Body language

- Looking uncomfortable was felt to be a sign that the patient was experiencing an issue that may not have been expressed, 'if they were grimacing or if they were holding their body in an unrelaxed position. So you know if people are clutching their abdomen or they're bent over'
- Appearing slumped was described by several participants as a possible indicator that the patient was too unwell to either do anything about their uncomfortable positioning or be aware of it. 'We went to see her and she'd kind of slipped down the bed and was looking a bit awkward' was the description of one patient whose NEWS2 score deteriorated later that day.

Behaviour

- New agitation in patients who were otherwise alert and oriented was described as a good screening tool- 'At that time they're there talking full sentences. A little bit agitated, but in the grand scheme of things you do a fresh set of obs on them and they're fine, the obs are, maybe slightly tachy I don't know 110-115 around that range. You do an ECG on them. You do an ECG, get a doctor to review it, go back 5-10 minutes later and they've gone off. Literally that quick'
- New lethargy in a patient who had previously been interacting with nursing and allied health professionals- 'She was, just different/ like really really different to her normal self.'

Patient Feeling unwell- New feeling of shortness of breath in the absence of deranged observations prior to overt deterioration. One participant described a renal patient she had been caring for for several shifts being, 'more breathless. But again you put that down to anxiety and being a little bit worked up. But obs stable, all fine and reassuring him….So I just documented it, made sure he was on the cardiac monitor, because he was, did his obs, checked his BMs gave him a bottle so I could do a urine dip. He didn't have any pain or anything. Checked he'd had all of his meds through the day. And that was about it. Made him a cup of tea…. the breathlessness got worse and he became more panicked and it became apparent that his obs then did start to deteriorate…Then all of a sudden he was in pulmonary oedema.' Feeling of impending doom was described by several participants and was often associated with agitation or what may have been viewed as confused behavior. In describing one patient who went on to arrest, a participant described, 'he just kept saying I need to speak with my wife. I need to speak with my wife. I'm going to die. I'm going to die. Which made me think, is he brewing something? You know. But this is highly unlike him. He's been absolutely fine my other night shifts and during the days. Just totally out of character'

Pain- More than expected or in the wrong place was described in both medical and surgical patients, where a patient had come in with a specific condition but had pain that was unlikely to be associated with it, or were post op with pain that was out of proportion to the procedure they had had or the time that had passed since the procedure , 'if someone couldn't roll post op because of pain and you'd given them quite a lot of pain relief, you'd usually be quite worried and taking them to scans to make sure they're not bleeding'

Reduced mobility in comparison with prior encounters with a patient was suggested by multiple participants as being an indicator of deterioration worthy of further investigation, 'So the fact they could walk yesterday and today they can't roll in the bed that doesn't always get as much recognition'

New Neurology not identified using the current AVPU assessment of conscious level was highlighted as an important early sign in several organ systems.

- **Change to speech** was the first sign noted by a participant whose patient was later identified as having had a stroke, 'she couldn't speak. And I knew I was chatting away with this lady last night, so I knew something was wrong'
- **New Weakness or coordination difficulties** were also identified by multiple participants as potential indicators of stroke, 'The way they're talking, you know,

yeh, level of consciousness, level of facial express, you know sometimes patients have facial drop'

- **Changes to vision** was highlighted as a sign of metabolic issues by one participant, 'He was pulled over by the police, they thought he'd been drinking, but his vision went when he was driving, and they breathalysed him and he was fine, and he ended up coming straight to us with a really high potassium for emergency dialysis'
- **Short-term memory issues** were described by one participant as triggering concern, 'not remembering something, like telling them we're going to be giving a blood transfusion, going back to flush the cannula and they haven't remembered. Whereas that wasn't something that was normal for them.'

Change in physical appearance was highlighted as part of the almost subconscious end of the bed assessment done during drug rounds and while collecting vital signs.

- **Skin colour** was reported as being used by every participant to assess patient condition with variations including pallor, jaundice and grey tone, 'He had a grey colour on his face, he was sweating, and he was mottled to the knees, so this was enough to convince me that he was hypoperfused and I was concerned about him.'
- **Skin feel**, particularly being clammy or cool, was also a universally reported sign of concern to participants
- **Breathing** pattern and use of accessory muscles was reported by the majority of participants as a useful sign not reflected in respiratory rate, but potentially an early sign of tiring or of underlying diagnosis. One participant described the following in a post op patient, 'Patient had a quite laboured respiratory rate. It wasn't fast enough to justify concern but it was laboured, he was using accessory muscles that he shouldn't be using'. As vital signs were normal the participant struggled to engage the parent team, even at consultant level, with the patient being transferred to critical care an hour later.
- **Erythema around a wound or joint was described as an important early sign, which if acted on, could prevent harm**, 'if there's wound leakage or redness or any sign of infection. But they don't score. There's nothing to score on the EWS but the patient could be going septic'. The participant went on to comment that in this situation a raised EWS was a late sign.

Requiring adjuncts to maintain homeostasis

> It was raised as a point of interest that although patients requiring oxygen to maintain a normal oxygen level score higher on NEWS2, other methods of supporting vital signs do not increase the score e.g. Barehugger to maintain temperature; IV fluid to maintain blood pressure. 'So the barehugger was the first sign, why has she dropped her temperature? Yes little old ladies could have a low temperature, I understand, to a certain extent but not to 35, 34. And it's summer not winter'

It should be noted when considering additional signs of deterioration that there are patient factors, which confound the ability of staff to apply end of the bed assessment as well as impacting the accuracy of the NEWS2 score in predicting risk of deterioration. For example, skin colour can affect ability to assess for pallor and jaundice, while it has also been shown to reduce the accuracy of oxygen saturation probes [132]

Communication between staff, patients and relatives

Communication with families was reported to be an important source of clinical information, providing context on baseline, and monitoring. There were several instances of investigations triggered by family members leading to intervention. For example, one participant in a surgical setting reported, 'it was just when the daughter came later on in the evening. It was pretty immediate. I think she spent 20 minutes in there and came out and said she's not right and I'm worried that she's had a stroke because she's had one before' in a patient who went on to be confirmed as having had a stroke. A second participant in gynaecology reported, 'I've had a young girl come in before and her dad has said 'something is not right, I just know, I know something is not right', in a patient who went on to have a heavy bleed.

Ease of communicating concerns to medical staff was reported to vary by department. For example, in perioperative care it was easy to get a senior review of patients of concern, 'I would find it very easy to get a consultant anaesthetist to come and review a patient very easily. But if that's perhaps maybe because they know that you have the skills to assess them properly and if a nurse is saying that they're worried there's something to be worried about and they'll come and have a look. But equally if there wasn't anything, they're more happy that you've raised something than not raising it.'

The general opinion across both surgical and medical specialties was that it was more difficult to get junior doctor, i.e. those classed as a registrar or below, to see patients without a high early warning score, 'Literally we have to pull the doctor away look, we have seen this patient. We have seen the difference in this patient. This is what she was when she came in and this is where she is now, even

though she's not scoring. "what's the EWS, What's the EWS?" Forget about the EWS. This is what she is. She is sick. You need to come and see her, discuss the case with the consultant. ' With participants being clear that this had caused harm, 'So we've had some catastrophic episodes when I've been in surgery of failing to escalate deteriorating patients. And what it stems from is the nursing staff were escalating but the medical staff hadn't reviewed.'

The experience of nursing staff was also highlighted as a component in ability to communicate clinical concerns. One participant said of more junior staff, 'I think they struggle to know what is urgent and important, so they mention everything as a safeguard and then you decide what's important and what's not.' An issue with confidence in escalating was also identified, '

Lack of continuity of staffing

Ability to see trajectory was highlighted as an issue by several participants when assessing a patient compared to baseline, 'it's not something we're good at, in monitoring a baseline, they don't put well they're oriented to time and place, they just put confusion, which is so broad'. But there were also incidences where small changes, which would not be picked up when doing a single shift, were noted, ' We went to see her and she'd kind of slipped down the bed and was looking a bit awkward. And um we said to her how you want me to tie your hair back as you do, and she said, no don't bother. And we just thought that's really uncharacteristic behaviour… Anyway, by late morning she'd had a cardiac arrest and gone off to ICU. And it was just bizarre, and we talked about this afterwards …When you're looking after someone that closely and that intimately for so long you pick up different things, and I don't know what it was but she didn't want her hair tying back'.

Time and confidence in specialty

Knowledge of what to expect in specialty area currently working in was highlighted as being an issue for doctors and nurses new to an area. For example, renal patients may have a very different baseline and trajectory to surgical patients. If they have limited experience in a specialty due to re-allocation they may not recognise early signs of deterioration, 'a patient having a new treatment where they'd be at high risk of things like cytokine release syndrome and they start rigoring, but they have haven't spiked a temperature and they're not actually scoring but you know something bad is going to happen soon, er so those types of patients. Or their temperature is creeping up, 37.8, their obs are ok at the moment but you know things are going to get worse. Just because you know their background'. Experience of an area was also described as adding context that could be reassuring, 'If you've got a patient with low BP in oncology It might not be so much of a stress, you'd recheck the blood pressure in 15 minutes. Whereas if you've got someone who's post op they're like, you need

fluids straight away. I think it depends on where you work, it's like the expected for that patient group'.

Transient workforce

The frequent rotation of nursing staff due to training and re-allocation for service needs, led to participants from some areas to report significant issues with lack of continuity in workforce and familiarity of systems and patient cohort. The movement of nursing staff around the hospital, particularly during COVID to support operationally stretched areas, was also identified as a source of misunderstanding of experience level. One participant who was a charge nurse stated, 'You get coined by your level of experience but your breadth goes down definitely. And talking to colleagues that's a big thing. Because we've moved round like we've never moved around before and the expectation has been very different'. Another who was a specialist nurse, with a very detailed but specific area of knowledge, stated, 'I've got a dark blue uniform on. If I'm sent to a ward or to A&E which I have been done in the last couple of years. If I get sent to A&E I'm an absolute fish out of water, I haven't got a clue. Even though I've had some recovery experience but that's very controlled, that's my place I was used to. But you walk around in dark blue and the expectation is that you know. So you might have a junior member of staff that's escalating concerns to you. But you might not be as confident in that specialty.'

The issue of limited experience due to rotation was also recognised with junior medical staff. One participant reported 'so it should be 2 F2s, we don't have any F1s, but generally it's just 1 because of staffing… and they've been brought in this new system, which I don't agree with, but they feel that it's better for their learning. So previously they'd be with us for a 4 month block on one ward and they'd take time out to go to pre-admission, to theatre, things like that. But what they do now is a week on each area'. This was felt to be less of an issue with more senior doctors, 'I think because you've got consultants stay fairly fixed in a place for a period they get to know the nursing staff. You know faces, you know people you know how long they've been there. Whereas on the ward you get people rotating a lot.' This was particularly seen in more specialist areas, where one participant reported escalating concerns to critical care outreach instead of hospital at night, 'because sometimes in renal a scared junior doctor is of no use to me. Within such a specialism. They've come on the ward crying before and we've had to sit down and counsel them for half an hour because they don't know what to do'

Changes to nursing education

Only one of the participants had been qualified for less than 5 years. The other sixteen reflected on significant changes in nursing education. There was a perceived reduction in scope of practice, 'We've lost the wider assessment. If I was assessing, back to my days assessing patients on the medical ward if I had someone come in with heart failure I'd be looking myself to see how much oedema they had an how far it came up as a standard'. But also a greater reliance on numbers, 'it's very much tick box exercises, and they don't think outside the box, they don't think I'm scoring this number, what does that mean?'

6.3.1.2 Environment

When describing the interaction between staff members and between staff and patients or their families, a significant emphasis was placed on the barriers created by the structures and environment in which patient care was being delivered.

6.3.1.2.1 Staffing

Nurse to patient ratio was raised by every participant as limiting ability to effectively monitor patients, 'one thing I did like about crit care nursing you are there you're doing everything one to one and you can do it properly. You then go back to the ward and you have 8 patients. Some of whom you won't see for hours'.

Skill mix was generally reported to have deteriorated over several years, worsened following COVID. This in turn had a knock on effect on nursing behaviour, 'I think it's been worse in the last 2 years. We lost a lot of experienced nurses. Because of staffing shortages we can't spend the time with newly qualified nurses so they don't feel confident to escalate their concerns to the doctors.' I've heard nursing staff when I've said, have you escalated this to the doctors? They say, 'oh well they don't listen to us' we need a band 6 nurse or a band 7 nurse to go and talk to the doctors because they lose that confidence.'

Presence of high requirement patient- such as confused, acutely unwell or end of life- leaving staffing for rest of ward short, 'recently we had quite a young person pass away, but the support the family needed placed quite a heavy burden on the nursing staff. Erm, so that can be really difficult when I was in that room and I've got 7 other patients'

6.3.1.2.2 Location

Distance from other areas covered by same team was reported to be an issue for two participants who worked in wards distant from the main clinical areas. This led to issues getting support from the

hospital at night and critical care outreach teams to review patients they were concerned about outside of the hours when their usual clinical teams were available on the wards, 'well during the day we've got doctors on the ward so before 5 it's easy. Whether they come to see or not is another thing. I don't find it easy to get hold of H@N doctors. We had one come on last night. We had 2 jobs he did one, I said I've got another one I've put through will you come and see her, he said yes, yes and left the ward. I didn't know where he was.', confirmed by a participant on a different ward, 'Getting junior doctors in the specialty, getting medical on call to see patients is so so hard, that's why from experience that if someone is going to deteriorate we need to pick it up really quickly.'

6.3.1.3 Tasks

Tasks were described both in terms of creating a barrier to recognising deterioration, for example where documentation was done badly or time taken writing notes detracted from patient care; or as a facilitator, where clear documentation allowed identification of trajectory or where bedside tasks allowed monitoring of softer patient cues merely through the act of spending time with them. 'I don't see the numbers, I see the patient, talk to the patient. When you're doing the obs, talk to the patient, have a little look, say oh have you been to the toilet? Have you eaten? And you get the full picture within that 5 minutes'

6.3.1.4 Tools

Nervecentre

The devices carried by staff to facilitate the use of Nervecentre in recording clinical and communicating it to others were described universally as an improvement from previous practices, allowing more information to be visible remotely, allowing staff to monitor patients remotely while managing other tasks, and allowing more efficient communication and triage between different members of the team. However this was seen as removing individual thought by some participants, 'I think that Nervecentre has potentially taken away some of that. Some of the clinical judgement. So they're not looking at the mouth, the tongue, does the patient look dry. They do tend to rely on what the obs are telling them on the screen. And I might just, sometimes you get an inkling, say if the patient's telling you they've got a bit of indigestion. I think I've got heartburn. If it's a lady in particular I'll just do an ECG anyway because they just present differently don't they.'

NEWS2

A recurring observation was that one of the tools designed to facilitate patient monitoring, the automated escalation of NEWS2 scores reaching a certain threshold, was perceived as a barrier to doctors acting on nurse concern. It was felt to encourage prioritisation of patients with a higher NEWS2 score over low scoring patients who had other features worrying the nursing staff, 'it's not even the SHOs fault because it's not their specialty it's not always what they're interested in. They're going off what is on the screen, which is true but they're not with them all of the time'

The NEWS 2 score itself was also seen as a barrier to escalation in certain situations. This was attributed to the fact that trajectory reflected in the way vital signs are recorded within certain parameters, for example increasing oxygen requirement does not alter the score, while large fluctuations in blood pressure can be masked by the width of the designated physiological range. Also, NEWS2 is a snapshot of vital signs at the moment they were taken and a patient can easily deteriorate before the next set are due.

Additional data available

There were two categories of routinely collected clinical information identified by study participants as potentially being useful additional tools to the current NEWS2 criteria in predicting patient stability. The first was abnormal blood tests suggestive of organ dysfunction such as high lactate, low blood sugar and impaired kidney or liver function. The second was bedside observations not currently included in NEWS2, including urine output, or lack thereof, and more specialist factors such as drain output, be it more than expected or the wrong fluid.

6.4 Discussion

This study identified a wide range of situations where nursing staff reported utilising factors independent of, and in addition to, vital signs in order to assess patient condition using features not currently identified using the NEWS2 score . The tension this creates and subsequent potential outcomes is shown in Table6-2.

Table 6-2 Nurse concern versus NEWS2 concern and potential consequences

	Nurse Concern Present	Nurse Concern Absent
NEWS2 Concern Present	High confidence patient needs clinical review and potential change to management to avoid harm	Patient could be deteriorating- Potential to avoid unnecessary intervention if nursing opinion based on validated assessment
NEWS2 Concern Absent	Patient could be deteriorating- risk of false reassurance by low NEWS2	High confidence patient is stable

In analysing the participant responses, a number of barriers to accurately developing and escalating nurse concern were identified. These can be mapped to the different elements of the PETT model described earlier, and potential mitigations considered based on the experiences of the nursing staff interviewed.

6.4.1.1 Barriers to effectively forming an accurate assessment of patient condition and escalating concerns

The first is barriers linked to people related factors. The transient nature of the workforce, whether through turnover, moving staff due to service need elsewhere or junior skill mix, was frequently mentioned as being a barrier to incorporating end of the bed assessment into management. This was felt to be because it created a lack of familiarity with the patient cohort, making it more difficult to ascertain what the expected baseline and trajectory would be in a particular group of patients. Participants also felt this contributed to a lack of consistency within the team and therefore ability to build up relationships between the different members of the multidisciplinary team. This needs to be considered when management are moving nurses into more specialist areas, the examples particularly identified in this study being renal and haematology. With possible mitigations being specialty specific adjustments to NEWS2 based on the characteristics of the cohort in question, where they are found to vary from the accepted normal currently used in NEWS2, and more frequent MDT interactions to improve communication and trust.

Lack of continuity of care in a particular patient journey, either through movement of nursing staff to different areas, or shift patterns, was also identified as a factor in reducing ability to identify and

highlight subtle changes. Staff felt that important soft cues, i.e. those without definite end points or numerical scales attached, could be seen in changes in behaviour from a previous day or shift before any evidence of deterioration could be seen in vital signs. They were also concerned that these cues could be missed where continuity of care was interrupted. This could be mitigated by a greater emphasis on the opinions of friends and family where patients are showing uncharacteristic behaviour. Highlighting level of engagement on a number of levels and changes from baseline.

Several participants in this study suggested that more junior staff have greater difficulty in raising concerns where they are at odds with the vital signs assessment of a patient, and, in their experience, lacked discernment in recognising and escalating softer cues. There was also feedback from nursing staff that concerns are often taken more seriously if raised by a more senior member of staff, such as a band 6 or above. Addressing this would require strategies to improve communication with the medical team to ensure nursing staff feel their concerns are likely to be taken seriously. A further potential change suggested by participants was changes to job plans at more senior levels to integrate more clinical time to ensure the experience of band 7 nurses and above is not lost and that more senior staff members are consistently available to support decision making at a junior level.

Environmental barriers to incorporating effective use of nurse concern identified in this study included the location of the patient, i.e. distance from other hospital resources such as specialty doctors, hospital at night and other wards, and the staff to patient ratio in that location. The distance from hospital resources was particularly felt to be an issue out of hours when ward-based clinicians go home and a relatively small team of junior doctors provide out of hours cover. In support of this, routing as a factor influencing task prioritisation has previously been identified with regards to decision making by junior doctors [133] particularly at more junior stages of training[134]. Possible ways to mitigate these factors include structuring of resources to support wards when planning service provision, particularly out of hours. Ensuring mandatory minimum staffing levels are observed allowing nursing staff the opportunity to use their judgement at the bedside. Providing more junior staff with training regarding task prioritisation is also likely to increase the chances that clinical factors, rather than ease of completing review, are taken into account.

The most frequently highlighted tool related barrier identified by participants was the removal of individual assessment in the prioritising of those metrics presented as part of NEWS2. Several participants also mentioned that the volume of scores that NEWS2 highlighted to the medical team left them little time to review patients with lower NEWS2 score where nurses raised concerns based solely on an end of the bed assessment. The snapshot nature of the score was also mentioned as providing false reassurance when vital signs may have a negative trajectory that isn't picked up and

change from minute to minute. Several of the nursing staff with experience in more specialised areas, such as renal medicine, identified issues relating to the universal nature of the score. That is due to the generic 'normal' ranges applied to the vital sign risk groupings that form the score it could fail to pick up nuances created by specialty. Participants stated that they had observed this to cause unnecessary tasks in stable patients whose baseline physiology meant they had a persistent high score, but also create false reassurance of stability in patients whose deterioration was masked by their underlying condition.

The final area of the PETT model relates to tasks and here it was the more senior nurses involved in the study who voiced frustration regarding barriers to care. Several felt that the current structure of paperwork lent itself more to a tick box exercise and less to an accurate reflection of the current clinical picture of the patient, particularly with relation to accurate documentation of mobility and communication, but also ability and interest relating to self-care. But also that the time taken to fill in the mandated assessments took them away from the patient care tasks which allowed them to monitor changes in their patients such as loss of interest, or loss of appetite. They felt it was important that nursing staff were able to be more involved in activities such as washing and nutrition and that more junior nurses and students should be educated regarding the opportunities that these tasks created for an all round evaluation of their patients.

6.4.1.2 Design interventions to accommodate patient concern

There are several ways in which the concept of bedside assessment and nursing concern could be developed and integrated in the systems which form the core of recognition and escalation of patient deterioration. the first is to incorporate it within the NEWS2 score either in a format similar to that used by the Australian 'Between the Flags' system as demonstrated in Table 6-3.

Table 6-3 Potential design intervention to create calling criteria used as a surrogate for high NEWS2 score where vital signs do not reflect the current nursing assessment

Amber Criteria- treat as NEWS2>=5 in terms of escalation and observation frequency	
Increasing Oxygen requirement	FiO2<40% to maintain target saturations
Abnormal Skin Colour	Pale/ Grey/ Jaundiced (new)
Neurological deficit	Changes to speech/ Communication/ New weakness/New pupillary defect/Hallucination
Agitation	In the absence of confusion
Reduced energy levels	Slumped body position, not interacting
Low urine output	Not PU'd for 6 hours or <0.5ml/kg for 6 hours if catheterised
Kidney function	New AKI on bloods
Unexpected drain output	More than expected or wrong fluid- e.g. blood or chyle
New/ increasing pain	Not consistent with current diagnosis- e.g. new calf pain in someone admitted with upper limb injury Not improving with treatment
Blood sugar	<4mmol/l without drop in conscious level
Patient feels unwell	Impending doom/ feels like they're going to die
Negative trajectory	Not getting better, unable to do things that were possible previously- e.g. physio
Temperature	Requiring external heat to maintain temperature e.g. warm air blanket
Red Criteria- treat as NEWS2>=7 in terms of escalation and obs frequency	
Increasing Oxygen requirement	FiO2 >40% to maintain target sats
Arterial oxygen	pO2 <10
Arterial CO2	pCO2>6
Blood gas pH	pH<7.2 or BE< -5
Lactate	>=4
Blood sugar	<4 with decreased conscious level (if unresponsive call 2222)
Not improving following clinical review	Amber criteria not reversed within 1 hour following review
Chest drain inserted for pneumothorax	Not swinging or bubbling

An alternative could be to create additional scoring elements to be used within the current NEWS structure, an approach consistent with the DENWIS group[127]. An example of a potential set of additional elements identified by participants in our study, and their associated scores, is shown in table 6-4. All of these features were suggested by participants in this study, with cut points suggested either by data analysis in the course of the studies described earlier in the thesis, with regards to urine output and increasing fraction of inspired oxygen, or from the literature with regards to blood gas abnormalities.

Table 6-4 Potential additional elements to be incorporated into NEWS based on the feedback from this qualitative study (left white), logistic regression analysis of vital signs data (shaded grey) and current clinical practice in the literature (yellow).

Test type	Observed abnormality	Cut point or criterion	Additional NEWS score
Bedside observations	Increasing Fio2 to maintain target sats	Current FiO2 <=40%	2
		Current FiO2 >40%	3
	Reduced urine output	Not PU'd for >6 hours or <0.5ml/kg for 6 hours if catheterised	2
		Not PU'd for >12 hours or <0.5ml/kg for 12 hours if catheterised	3
	Unexpected drain output	Too high or wrong fluid (blood/chyle/bile) Chest drain inserted for pneumothorax not swinging or bubbling	3
Patient appearance	Abnormal skin	Pale/ Grey/ Clammy/ Cool/ Newly Jaundiced	3
	New Neurological deficit	Speech/ Communication/ Weakness/ Pupillary defect/ Hallucination	3
	General deterioration during admission	NEW: Slumped body position Not interacting Unable to do physio Reduced mobility	
	Agitation in absence of confusion		3
Point of Care	Blood gas abnormalities	Arterial pCO2 >6.0 Venous pCO2 >6.5	3
		Arterial pO2 <10 if target sats 94-98 Arterial pO2 <7.5 if target sats 88-92	3
		Lactate >=4	3
		pH <7.2 or BE < -5	3
	Blood sugar	<4 no neurological deficit	3
		<4 with decreased conscious level	Call MET

Additional support	Requiring external warming to maintain temperature	e.g. needing warmed air blanked or warmed fluids	2
Other	New patient concern (as opposed to consistently anxious patient)	Feeling of impending doom/ Like going to die	3

6.4.2 Conclusion

The issues relating to lack of resources in terms of staffing numbers identified by participants in this study have been well described elsewhere as contributing to morbidity and mortality and remain largely outside the scope of individual organisations to address while staffing levels nationwide remain at their current levels.

However, several points were identified which could be used to design systems that support nursing staff in delivering safer patient care in a way they are uniquely placed to contribute. Utilising these data to design future systems that utilise signals outside of the current scope of the NEWS score creates the opportunity to empower and re-engage nursing staff in day to day patient care and provide a safer environment for patients. This would in turn have the potential to improve relationships between nursing and medical staff and reduce the number of unnecessary escalations to out of hours teams.

7 Discussion

The review of the existing literature surrounding early warning scores, their origins, development and their use in a hospital setting detailed in the first chapter of this thesis highlighted their increasingly prominent role at the centre of patient care. From simple vital signs based scoring systems based on expert opinion to aggregate weighted scores derived from outcomes-linked vital signs data, they have also become more complex over the 30 years since vital signs based triggers were first postulated. In the NHS, the National Early Warning Score version 2 (NEWS2) has become the score mandated for use in order to create a unified approach allowing for benchmarking of performance in relation to outcomes. However, the simplicity of using a single scoring system makes several assumptions regarding the inpatient population that had yet to be explored in the literature.

The overarching question driving this thesis of 'How effective are early warning scores in highlighting deterioration across the hospital?', cannot be looked at as a statistical problem in isolation. The NHS has finite resources in terms of medical and nursing staff to respond to clinical alerts. This means there needs to be a balance between sensitivity in detecting possible physiological decline and workload generated by observations reaching the set threshold for escalation and increased frequency of observations where no change in management is required.

7.1 Balance between workload and score sensitivity

Early warning scores have been developed based on statistical analysis of vital signs observations, with the focus on statistical discrimination of ability to predict outcome. However, no prospective studies have examined the impact of introducing a specific score as a stand-alone measure, i.e. with no other additions into the system. There were no prospective studies of any kind examining the impact of the National Early Warning Score before its initial release, nor prior to it becoming mandated by NHS England when version 2 was released. Therefore the systems impact of introducing a score with high sensitivity to a population with diverse physiology was poorly understood.

Because the major focus of early warning scores has been creating a safety net to avoid missing patients at risk of deterioration, sensitivity has always been prioritised in the research alongside area under the receiver operating characteristic curve. Sensitivity in predicting outcome is important for identifying patients with potentially deranged physiology, however identifying that someone is potentially more unstable does not necessarily mean there is anything that could be done to change clinical trajectory.

A study performed using change in management rather than outcome of death or ICU admission showed that less than 2% of observations in the 1100 respiratory admissions examined were followed by an intervention upon review of notes and drug charts [135]. This study was undertaken in the same respiratory department during the same period covered by this study and highlights an interesting point that of the 28% of observation sets reaching the scoring threshold for escalation to a junior doctor, and 11% to the registrar, only a small fraction is likely to have resulted in a change in management.

In Nottingham, the switch to NEWS2 led to a significant increase in demand on both the nursing and medical workforce due to increase in scores reaching threshold for both increase in frequency of taking vital signs and escalation. The Nottingham University Hospitals NHS Trust has a system of automatic escalation. Therefore unless there is an active decision by the nurse in charge not to escalate a set of observations reaching the threshold set by the NEWS2 protocol, it will automatically be referred to the appropriate person. This allows scores reaching threshold for escalation to be used as a reasonable predictor of demand, with recorded escalations before and after the introduction of NEWS2 reflecting this. A wide variation in the NNE values for escalation to the registrar, and therefore workload, was seen across the cohorts studied in chapter 4, with no clear relationship to mortality rate. Highlighting a question regarding the efficiency and effectiveness of the score in certain cohorts.

There is consensus that there is a limit to the processing capacity an individual can handle at any one time performing a particular task, and if overloaded beyond this capacity then performance reduces and potential for errors increase [105]. Therefore the balance between not escalating patients who need intervention and swamping nursing and medical staff with tasks relating to clinically stable patients is one that needs to be carefully weighed up.

7.2 Ways of improving NEWS using different score patterns

The hypothesis that patterns could play a role in improving the discrimination of NEWS2 in predicting outcome came from two sources in the early stages of setting the questions to be explored during the quantitative analysis carried out in this PhD. The first source was the literature review, where several studies examined mortality based on early trends in early warning score at admission [136] and trajectory of individual vital signs [55, 56, 137], but looked at trends in NEWS2 throughout admission. The second source was staff, who reported that whenever they're given a set

of observations they look for trend. Is this baseline for that patient? Is there a clear downwards or upwards trend?

This is distinct from more complex modelling of NEWS2 which could not currently be rolled out in many NHS trusts due to the lack of digital maturity identified as part of the freedom of information request detailed in chapter 2. One of the foci in this project has been to generate data that can be applied in any setting now. Several straightforward temporal pattern variables were therefore explored with the aim of potential implementation if any were found to add significantly to the risk prediction of the current NEWS2 score and protocol. Of those tested, maximum score in the preceding 24 hours was the most promising pattern variable and has the advantage of being usable in paper-based systems. In doing this, the importance of where escalation thresholds are set with relation to performance of the score itself and potential impact on workload as defined by NNE was further highlighted.

In the future wearable devices may allow physiological variations in vital sign values to be modelled more accurately and therefore negative trends to be identified earlier in clinical trajectory, potentially improving the sensitivity of these approaches. In addition, the more widespread adoption of integrated digital systems will allow the more complex AI interventions currently being developed, linking not just patterns but also interventions and response to them through linkage to electronic prescribing platforms in real time.

7.3 Ways to improve NEWS using different factors- (age criteria/diagnostic criteria, nurse concern)

The clinical judgement of experience nurses regarding the clinical status of their patients is well recognised in the clinical setting. However, for nursing concern to be used effectively it needs to be evidence-based and have clear parameters. The nurse concern study defined factors feeding into overall end of the bed assessment that participants felt could be integrated into scoring systems to improve discrimination. These included bedside factors based on appearance, and bedside measurements not currently included in NEWS2 such as urine output, drain output, change in inspired oxygen or use of an adjunct to maintain observations within the normal parameters, for example- use of iv fluid to maintain blood pressure or a barehugger to maintain temperature. The 'Between the Flags' system in Australia uses many of the factors mentioned in the nurse concern study, including end of the bed concern, blood sugar, lactate, urine output, as part of their amber and red criteria for triggering clinical review or medical emergency team activation. In addition, the studies reported here have demonstrated that demographic and diagnostic patient characteristics can provide additional information regarding risk prediction.

Leaving aside the many factors raised as barriers to or facilitators in escalation, it is important to highlight that all nursing staff felt that it was easier to gain a clinical review of a patient they were concerned about when their opinion matched the NEWS2 score. This issue appeared to be multifactorial, and related to several elements explored using the PETT model as described in Chapter 5. Therefore to have an impact, nurse concern needs be integrated into the score itself to give the escalations it triggers equal weighting to vital signs related triggers.

The use of patient and relative concern is being explored in other centres within the NHS. With an initiative called 'call 4 concern'[138] using concern to trigger a critical care outreach review reporting patients identified earlier in their clinical course than would have been the case relying on NEWS2 and therefore preventing further deterioration [139].

7.4 What the future holds

One common National Early Warning Score framework used in every hospital across the NHS is the only pragmatic way to move forward from this point. This is for all of the reasons that the development group and the Royal College of Physicians put forward, i.e. ability to benchmark care, sharing best practice and familiarity for staff moving hospital. But that does not mean applying the same rigid criteria for every patient. As demonstrated in this thesis a 25 year old surgical patient does not carry the same risk of mortality within the next 24 hours as an 85 year old medical patient with the same vital sign score. Nor do they have the same trajectory. There is also the question of safety. The introduction of scale 2 for oxygen saturations demonstrates that there are areas where concessions need to be made to disease processes where the targets applied to patients without chronic disease risk harm through chasing inappropriate targets. In addition creating unnecessary workload and intervention, while diverting resources away from patients who are deteriorating and contributing alert fatigue.

The findings set out in this thesis highlight what can, and should, be done to improve recognition of deterioration in the immediate future. With improvements that can be introduced into current systems, such as making evidence-based adjustments for patient groups that clearly have different characteristic, whether this be related to age or diagnosis. The qualitative study focussing on nurse concern has allowed us to postulate criteria which could be used in addition to the vital signs currently employed by NEWS2. Going forward, Nurse concern needs to be formalised and introduced with prospective studies determining which of the methods described in chapter 6 would have the most beneficial impact on prevention of deterioration, not just prediction of it. This could be done safely through adding nurse concern into systems in a way where they are initially recorded in the background as part of vital signs documentation without being acted on, before moving into a

controlled trial of full usage as part of the assessment of patient condition. With all future work including the specific point that has been missing from previous studies- what proportion of patients highlighted need intervention to prevent deterioration.

As hospitals are able to develop greater digital maturity, systems reliant on more complex patterns will become more possible, with several teams already working on dynamic scores as a means to improving prediction. The accuracy and speed of recognition of systems using this approach will also be impacted by the inevitable future introduction of wearable monitoring, allowing minute to minute assessment of patients outside of the critical care setting and therefore a greater understanding of what constitutes physiological versus pathophysiological change.

In all of this, detection of risk only provides half of the story. The focus on workload throughout this thesis has been important as a recognition that the NHS has limited resources. To make the best use of them we need to understand not just the most effective way of identifying patients who are at risk of decline, but also those in whom an intervention would change the outcome, and what that intervention should be. The relative merits of outreach nurses versus on call junior doctors versus a full on medical emergency team need to be better understood and evidenced to allow the correct response to each patient and situation to be made. In the future, if you are a 50 year old with chronic kidney disease and sepsis in a hospital in Newcastle that should be detected as quickly and responded to in the same way as a hospital in Southampton. It should be the patient and situation that determines the action, not their geographical location.

7.4.1 Contributions to previous knowledge

First and foremost, the impact of introducing NEWS2 to a system which previously had a mature electronic observations system with automated referrals has not previously been demonstrated. This included sizeable increases in numbers of vital signs observations sets collected and number of escalations to different clinician groups per day without any increase in ward staffing or out of hours clinicians.

Previous studies have demonstrated the ability of NEWS to predict outcomes in certain specific populations. This thesis takes that idea further and is to our knowledge the most comprehensive breakdown of the ability of NEWS2 to predict outcome of death within 24 hours by age and diagnostic group. The statistical impact of readily available information including age, gender, presence of previous admissions and timing of admission on outcomes including mortality and length of stay in medical and surgical populations is also novel as far as we are aware within the

literature. These data are the first step in the development of personalised scoring systems based on patient characteristics.

The use of pattern of individual vital signs observations has previously been explored. However, additive predictive value of a pattern variable of NEWS2 has not previously been assessed. This represents a straightforward and universally applicable method of improving prediction through adding a temporal component to the current snapshot of vital signs.

Indicators of nurse concern have previously been postulated following systematic review of many smaller studies. However this thesis is the first time this has been looked at prospectively using semi-structured interviews to explore the factors used by nursing staff to assess patient condition and guide understanding of which elements are important when using NEWS2 in an NHS setting.

7.5 Conclusion

The National Early Warning Score in its current format does not go far enough towards identifying which patients need intervention where patient characteristics mean their baseline physiology, and reaction to pathological stimulus, is divergent from the population norms represented within the score. This thesis provides clear direction as to where improvements could and should be made in the short, medium and long term.

In all of this the most important thing going forward is to be able to access the vital data to allow us to power this work, and to collaborate with others. The challenges in accessing the data required for the studies described here were not unique to these projects. The barriers to working on anonymised data, even by NHS employees performing studies sanctioned by the HRA is significant. But if we can overcome this then the power of being able to work with other centres will multiply the power of the data and the significance of the findings. We need trusted research environments allowing the data of multiple organisations to be analysed in a secure and timely manner to the genuine benefit of patients. We need to move away from protecting our individual silos and build trusted research partnerships that allow us to focus the collective knowledge and evidence on truly moving patient care into the digital age.

References

1. Digital, N., *Hopsital Admitted Patient Care and Adult Critical Care Activity 2017-2018*, N. Digital, Editor. 2018.

2. Mahon, A., et al., *The implementation and impact of Hospital at Night pilot projects. An evaluation report*, D.o. Health, Editor. 2005, Department of Heakth.

3. Digital, N., *Summary Hospital-Level Mortality Indicator (SHMI)- Deatgs associated with hospitalisationm England, April 2017- March 2018*, in *Summary Hospital-level Indicator-deaths associated with hospitalisation*. 2018.

4. ICNARC, *Case Mix Programme Summary Statistics 2017/18*. 2018.

5. National Cardiac Arrest, A., *Key Summary Statistics 2017-18*. 2018.

6. Kause, J., et al., *A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom--the ACADEMIA study.* Resuscitation, 2004. **62**(3): p. 275-82.

7. Buist, M., et al., *Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study.* Resuscitation, 2004. **62**(2): p. 137-41.

8. Hillman, K.M., et al., *Duration of life-threatening antecedents prior to intensive care admission.* Intensive Care Med, 2002. **28**(11): p. 1629-34.

9. Hillman, K.M., et al., *Antecedents to hospital deaths.* Intern Med J, 2001. **31**(6): p. 343-8.

10. Goldhill, D.R., S.A. White, and A. Sumner, *Physiological values and procedures in the 24 h before ICU admission from the ward.* Anaesthesia, 1999. **54**(6): p. 529-34.

11. McQuillan, P., et al., *Confidential inquiry into quality of care before admission to intensive care.* BMJ, 1998. **316**.

12. Franklin, C. and J. Mathew, *Developing strategies to prevent inhospital cardiac arrest: analyzing responses of physicians and nurses in the hours before the event.* Crit Care Med, 1994. **22**(2): p. 244-7.

13. Schein, R.M., et al., *Clinical antecedents to in-hospital cardiopulmonary arrest.* Chest, 1990. **98**(6): p. 1388-92.

14. Andersen, L.W., et al., *The prevalence and significance of abnormal vital signs prior to in-hospital cardiac arrest.* Resuscitation, 2016. **98**: p. 112-7.

15. Sax, F.L. and M.E. Charlson, *Medical patients at high risk for catastrophic deterioration.* Crit Care Med, 1987. **15**(5): p. 510-5.

16. Husband, A., et al., *The epidemiology of respiratory arrests in a teaching hospital.* Resuscitation, 2014. **85**(3): p. 364-8.

17. DeVita, M.A., et al., *"Identifying the hospitalised patient in crisis"--a consensus conference on the afferent limb of rapid response systems.* Resuscitation, 2010. **81**(4): p. 375-82.

18. McBride, J., et al., *Long-term effect of introducing an early warning score on respiratory rate charting on general wards.* Resuscitation, 2005. **65**(1): p. 41-4.

19. Chen, J., et al., *The impact of introducing medical emergency team system on the documentations of vital signs.* Resuscitation, 2009. **80**(1): p. 35-43.

20. Cahill, H., et al., *Introduction of a new observation chart and education programme is associated with higher rates of vital-sign ascertainment in hospital wards.* BMJ Qual Saf, 2011. **20**(9): p. 791-6.

21. Preece, M.H., et al., *Supporting the detection of patient deterioration: observation chart design affects the recognition of abnormal vital signs.* Resuscitation, 2012. **83**(9): p. 1111-8.

22. Christofidis, M.J., et al., *Observation chart design features affect the detection of patient deterioration: a systematic experimental evaluation.* J Adv Nurs, 2016. **72**(1): p. 158-72.

23. Arora, M., et al., *Evaluation of CoViSTA - an automated vital sign documentation system - in an inpatient hospital setting.* AMIA Annu Symp Proc, 2005: p. 885.

24. Elliott, D., et al., *User compliance with documenting on a track and trigger-based observation and response chart: a two-phase multi-site audit study.* J Adv Nurs, 2017. **73**(12): p. 2892-2902.

25. McCambridge, J., J. Witton, and D.R. Elbourne, *Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects.* J Clin Epidemiol, 2014. **67**(3): p. 267-77.

26. Gearing, P., et al., *Enhancing patient safety through electronic medical record documentation of vital signs.* J Healthc Inf Manag, 2006. **20**(4): p. 40-5.

27. Prytherch, D.R., et al., *Calculating early warning scores--a classroom comparison of pen and paper and hand-held computer methods.* Resuscitation, 2006. **70**(2): p. 173-8.

28. Foley, C. and M. Dowling, *How do nurses use the early warning score in their practice? A case study from an acute medical unit.* J Clin Nurs, 2019. **28**(7-8): p. 1183-1192.

29. Petersen, J.A., L.S. Rasmussen, and S. Rydahl-Hansen, *Barriers and facilitating factors related to use of early warning score among acute care nurses: a qualitative study.* BMC Emerg Med, 2017. **17**(1): p. 36.

30. Recio-Saucedo, A., et al., *Relationships between healthcare staff characteristics and the conduct of vital signs observations at night: Results of a survey and factor analysis.* Nurs Open, 2018. **5**(4): p. 621-633.

31. Hope, J., et al., *A fundamental conflict of care: Nurses' accounts of balancing patients' sleep with taking vital sign observations at night.* J Clin Nurs, 2018. **27**(9-10): p. 1860-1871.

32. Griffiths, P., et al., *Vital signs monitoring in hospitals at night.* Nurs Times, 2015. **111**(36-37): p. 16-7.

33. Gordon, C.F. and D.J. Beckett, *Significant deficiencies in the overnight use of a Standardised Early Warning Scoring system in a teaching hospital.* Scott Med J, 2011. **56**(1): p. 15-8.

34. Hands, C., et al., *Patterns in the recording of vital signs and early warning scores: compliance with a clinical escalation protocol.* BMJ Qual Saf, 2013. **22**(9): p. 719-26.

35. Yiu, C.J., et al., *Into the night: factors affecting response to abnormal Early Warning Scores out-of-hours and implications for service improvement.* Acute Med, 2014. **13**(2): p. 56-60.

36. Weenk, M., et al., *Continuous Monitoring of Vital Signs Using Wearable Devices on the General Ward: Pilot Study.* JMIR Mhealth Uhealth, 2017. **5**(7): p. e91.

37. Prgomet, M., et al., *Vital signs monitoring on general wards: clinical staff perceptions of current practices and the planned introduction of continuous monitoring technology.* Int J Qual Health Care, 2016. **28**(4): p. 515-21.

38. Weenk, M., et al., *Wireless and continuous monitoring of vital signs in patients at the general ward.* Resuscitation, 2019. **136**: p. 47-53.

39. Downey, C., et al., *Reliability of a wearable wireless patch for continuous remote monitoring of vital signs in patients recovering from major surgery: a clinical validation study from the TRaCINg trial.* BMJ Open, 2019. **9**(8): p. e031150.

40. Van Velthoven, M.H., et al., *ChroniSense National Early Warning Score Study (CHESS): a wearable wrist device to measure vital signs in hospitalised patients-protocol and study design.* BMJ Open, 2019. **9**(9): p. e028219.

41. Hourihan, F., et al., *The Medical Emergency Team: A new strategy to identify and intervene in high-risk patients.* Clinical Intensive Care, 1995. **6**(6): p. 269-272.

42. Morgan, R.W., et al., *An early warning scoring system for detecting developing critical illness.* Clinical Intensive Care, 1997. **8**(2): p. 100.

43. Fernando, S.M., et al., *Prognostic accuracy of the Hamilton Early Warning Score (HEWS) and the National Early Warning Score 2 (NEWS2) among hospitalized patients assessed by a rapid response team.* Crit Care, 2019. **23**(1): p. 60.

44. Churpek, M.M., et al., *Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*.* Crit Care Med, 2014. **42**(4): p. 841-8.

45. Prytherch, D.R., et al., *ViEWS--Towards a national early warning score for detecting adult inpatient deterioration.* Resuscitation, 2010. **81**(8): p. 932-7.

46. Cuthbertson, B.H., M. Boroujerdi, and G. Prescott, *The use of combined physiological parameters in the early recognition of the deteriorating acute medical patient.* J R Coll Physicians Edinb, 2010. **40**(1): p. 19-25.

47. Smith, G.B., et al., *Review and performance evaluation of aggregate weighted 'track and trigger' systems.* Resuscitation, 2008. **77**(2): p. 170-9.

48. Duckitt, R.W., et al., *Worthing physiological scoring system: derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study.* Br J Anaesth, 2007. **98**(6): p. 769-74.

49. Subbe, C.P., et al., *Validation of a modified Early Warning Score in medical admissions.* QJM, 2001. **94**(10): p. 521-6.

50. Smith, G.B., *Have we found the perfect early warning score? A view of ViEWS.* Resuscitation, 2013. **84**(6): p. 707-8.

51. Kovacs, C., et al., *Comparison of the National Early Warning Score in non-elective medical and surgical patients.* Br J Surg, 2016. **103**(10): p. 1385-93.

52. Ghosh, E., et al., *Early Deterioration Indicator: Data-driven approach to detecting deterioration in general ward.* Resuscitation, 2018. **122**: p. 99-105.

53. Watkinson, P.J., et al., *Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data.* Resuscitation, 2018. **129**: p. 55-60.

54. Pimentel, M.A.F., et al., *A comparison of the ability of the National Early Warning Score and the National Early Warning Score 2 to identify patients at risk of in-hospital mortality: A multi-centre database study.* Resuscitation, 2019. **134**: p. 147-156.

55. Churpek, M.M., R. Adhikari, and D.P. Edelson, *The value of vital sign trends for detecting clinical deterioration on the wards.* Resuscitation, 2016. **102**: p. 1-5.

56. Brekke, I.J., et al., *The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review.* PLoS One, 2019. **14**(1): p. e0210875.

57. Smith, G.B., et al., *The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death.* Resuscitation, 2013. **84**(4): p. 465-70.

58. Hodgson, L.E., et al., *A validation of the National Early Warning Score to predict outcome in patients with COPD exacerbation.* Thorax, 2017. **72**(1): p. 23-30.

59. Le Lagadec, M.D., T. Dwyer, and M. Browne, *The efficacy of twelve early warning systems for potential use in regional medical facilities in Queensland, Australia.* Aust Crit Care, 2019.

60. Cretikos, M., et al., *The objective medical emergency team activation criteria: a case-control study.* Resuscitation, 2007. **73**(1): p. 62-72.

61. Smith, G.B., et al., *A review, and performance evaluation, of single-parameter "track and trigger" systems.* Resuscitation, 2008. **79**(1): p. 11-21.

62. Churpek, M.M., T.C. Yuen, and D.P. Edelson, *Risk stratification of hospitalized patients on the wards.* Chest, 2013. **143**(6): p. 1758-1765.

63. Green, M., et al., *Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients.* Resuscitation, 2018. **123**: p. 86-91.

64. Smith, G.B., et al., *Re: cross-sectional audit on the relevance of Elevated National Early Warning Score in medical patients at a Model 2 hospital in Ireland.* Ir J Med Sci, 2016. **185**(1): p. 267-8.

65. Haegdorens, F., et al., *The introduction of a rapid response system in acute hospitals: A pragmatic stepped wedge cluster randomised controlled trial.* Resuscitation, 2018. **129**: p. 127-134.

66. Haegdorens, F., et al., *An intervention including the national early warning score improves patient monitoring practice and reduces mortality: A cluster randomized controlled trial.* J Adv Nurs, 2019.

67. Parr, M., et al., *The Medical Emergency Team: 12 Month analysis of reasons for activation, immediate outcome and not-for-resuscitation orders.* Resuscitation, 2001. **50**(1): p. 39-44.

68. Chen, J., et al., *Triggers for emergency team activation: a multicenter assessment.* J Crit Care, 2010. **25**(2): p. 359 e1-7.

69. Romero-Brufau, S., et al., *The fifth vital sign? Nurse worry predicts inpatient deterioration within 24 hours.* Journal of American Informatics Association, 2019.

70. Andrews, T. and H. Waterman, *Packaging: a grounded theory of how to report physiological deterioration effectively.* J Adv Nurs, 2005. **52**(5): p. 473-81.

71. Douw, G., et al., *Nurses' worry or concern and early recognition of deteriorating patients on general wards in acute care hospitals: a systematic review.* Crit Care, 2015. **19**: p. 230.

72. Royal College of Physicians of, L., *National Early Warning Score (NEWS); Standardising the assessment of acute illness severity in the NHS*. 2012.

73. Pedersen, N.E., et al., *Modifications of the National Early Warning Score for patients with chronic respiratory disease.* Acta Anaesthesiol Scand, 2018. **62**(2): p. 242-252.

74. Eccles, S.R., et al., *CREWS: improving specificity whilst maintaining sensitivity of the National Early Warning Score in patients with chronic hypoxaemia.* Resuscitation, 2014. **85**(1): p. 109-11.

75. Royal College of Physicians of, L., *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS*. 2017.

76. Austin, M.A., et al., *Effect of high flow oxygen on mortality in chronic obstructive pulmonary disease patients in prehospital setting: randomised controlled trial.* BMJ, 2010. **341**: p. c5462.

77. Echevarria, C., J. Steer, and S.C. Bourke, *Comparison of early warning scores in patients with COPD exacerbation: DECAF and NEWS score.* Thorax, 2019. **74**(10): p. 941-946.

78. Zealand, T.S.o.A.a.N., *Oxygen guidelines for acute oxygen use in adults; Clinical Practice Guideline- "Swimming between the flags"*. 2015.

79. Chu, D.K., et al., *Mortality and morbidity in acutely ill adults treated with liberal versus conservative oxygen therapy (IOTA): a systematic review and meta-analysis.* Lancet, 2018. **391**(10131): p. 1693-1705.

80. Improvement, N., *The adult patient who is deteriorating: sharing learning from literature, incident reports and root cause analysis investigations*. 2016.

81. Dalton, M., et al., *Factors that influence nurses' assessment of patient acuity and response to acute deterioration.* Br J Nurs, 2018. **27**(4): p. 212-218.

82. Kalliokoski, J., et al., *Insight into hospital ward nurses' concerns about patient health and the corresponding Medical Emergency Team nurse response.* Intensive Crit Care Nurs, 2019. **53**: p. 100-108.

83. Chua, W.L., et al., *A call for better doctor-nurse collaboration: A qualitative study of the experiences of junior doctors and nurses in escalating care for deteriorating ward patients.* Aust Crit Care, 2019.

84. Mohammmed Iddrisu, S., et al., *Nurses' role in recognising and responding to clinical deterioration in surgical patients.* J Clin Nurs, 2018. **27**(9-10): p. 1920-1930.

85. Metcalfe, L., S. McNally, and S.M. Smith, *A review of inpatient ward location and the relationship to Medical Emergency Team calls.* Int Emerg Nurs, 2017. **31**: p. 52-57.

86. Subbe, C.P., B. Duller, and R. Bellomo, *Effect of an automated notification system for deteriorating ward patients on clinical outcomes.* Crit Care, 2017. **21**(1): p. 52.

87. Patel, S., S.A. Gillon, and D.A. Jones, *Rapid response systems: recognition and rescue of the deteriorating hospital patient.* Br J Hosp Med (Lond), 2017. **78**(3): p. 143-148.

88. Tillmann, B.W., et al., *The impact of delayed critical care outreach team activation on in-hospital mortality and other patient outcomes: a historical cohort study.* Can J Anaesth, 2018. **65**(11): p. 1210-1217.

89. Ball, C., M. Kirkby, and S. Williams, *Effect of the critical care outreach team on patient survival to discharge from hospital and readmission to critical care: non-randomised population based study.* BMJ, 2003. **327**(7422): p. 1014.

90. Jeddian, A., et al., *Evaluation of a critical care outreach service in a middle-income country: A stepped wedge cluster randomized trial and nested qualitative study.* J Crit Care, 2016. **36**: p. 212-217.

91. Hillman, K., et al., *Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial.* Lancet, 2005. **365**(9477): p. 2091-7.

92. Aneman, A., et al., *Characteristics and outcomes of patients admitted to ICU following activation of the medical emergency team: impact of introducing a two-tier response system.* Crit Care Med, 2015. **43**(4): p. 765-73.

93. Pain, C., et al., *Between the flags: implementing a safety-net system at scale to recognise and manage deteriorating patients in the New South Wales Public Health System.* Int J Qual Health Care, 2017. **29**(1): p. 130-136.

94. Royal College of Physicians of, L., *National Early Warning Score (NEWS) 2*, in *Standardising the assessment of acute-illness severity in the NHS*. 2018, Royal College of Physicians: Royal College of Physicians Website.

95. Pankhurst, T., et al., *Evaluation of NEWS2 response thresholds in a retrospective observational study from a UK acute hospital.* BMJ Open, 2022. **12**(2): p. e054027.

96. Romero-Brufau, S., et al., *Why the C-statistic is not informative to evaluate early warning scores and what metrics to use.* Crit Care, 2015. **19**: p. 285.

97. Kipnis, P., et al., *Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU.* J Biomed Inform, 2016. **64**: p. 10-19.

98. Collins, J.A., et al., *Relating oxygen partial pressure, saturation and content: the haemoglobin-oxygen dissociation curve.* Breathe (Sheff), 2015. **11**(3): p. 194-201.

99. Saito, T. and M. Rehmsmeier, *The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets.* PLoS One, 2015. **10**(3): p. e0118432.

100. Fang, A.H.S., W.T. Lim, and T. Balakrishnan, *Early warning score validation methodologies and performance metrics: a systematic review.* BMC Med Inform Decis Mak, 2020. **20**(1): p. 111.

101. Jarvis, S.W., et al., *Are observation selection methods important when comparing early warning score performance?* Resuscitation, 2015. **90**: p. 1-6.

102. Malycha, J., et al., *The effect of fractional inspired oxygen concentration on early warning score performance: A database analysis.* Resuscitation, 2019. **139**: p. 192-199.

103. Hydes, T.J., et al., *National Early Warning Score Accurately Discriminates the Risk of Serious Adverse Events in Patients With Liver Disease.* Clin Gastroenterol Hepatol, 2018. **16**(10): p. 1657-1666 e10.

104. Badriyah, T., et al., *Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS).* Resuscitation, 2014. **85**(3): p. 418-23.

105. Moray, N., *Models and Measures of Mental Workload*, in *Mental Workload: Its Theory and Measurement*, N. Moray, Editor. 1979, Springer US: Boston, MA. p. 13-21.

106. Perez, I., et al., *Out of hours workload management: Bayesian inference for decision support in secondary care.* Artif Intell Med, 2016. **73**: p. 34-44.

107. Forster, S., et al., *Investigating the discriminative value of Early Warning Scores in patients with respiratory disease using a retrospective cohort analysis of admissions to Nottingham University Hospitals Trust over a 2-year period.* BMJ Open, 2018. **8**(7): p. e020269.

108. Spangfors, M., M. Molt, and K. Samuelson, *National Early Warning Score: A survey of registered nurses' perceptions, experiences and barriers.* J Clin Nurs, 2020. **29**(7-8): p. 1187-1194.

109. Harrell, F.E., Jr., *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* 2001, New York: Springer

110. Digital, N. *Summary Hospital Mortality rate Indicator: Diagnosis group breakdown*. 2022; Available from: https://app.powerbi.com/view?r=eyJrIjoiMjNkM2JiNmQtMjc5NS00YjE2LTljNDgtZDZmNmYzZDM1ZTFhIiwidCI6IjUwZjYwNzFmLWJiZmUtNDAxYS04ODAzLTY3Mzc0OGU2MjllMiIsImMiOjh9 .

111. Chester, J.G. and J.L. Rudolph, *Vital signs in older patients: age-related changes.* J Am Med Dir Assoc, 2011. **12**(5): p. 337-43.

112. Nissen, S.K., et al., *The Impact of Age on Predictive Performance of National Early Warning Score at Arrival to Emergency Departments: Development and External Validation.* Ann Emerg Med, 2022. **79**(4): p. 354-363.

113. Smith, G.B., et al., *Should age be included as a component of track and trigger systems used to identify sick adult patients?* Resuscitation, 2008. **78**(2): p. 109-15.

114. Shamout, F., et al., *Early warning score adjusted for age to predict the composite outcome of mortality, cardiac arrest or unplanned intensive care unit admission using observational vital-sign data: a multicentre development and validation.* BMJ Open, 2019. **9**(11): p. e033301.

115. Forster, S., et al., *Predicting outcome in acute respiratory admissions using patterns of National Early Warning Scores.* Clin Med (Lond), 2022. **22**(5): p. 409-415.

116. Organisation, W.H. *International Statistical Classification of Diseases and Related Health Problems 10th Revision version XXII*. 2019; Available from: https://icd.who.int/browse10/2019/en#/XXII.

117. Forster, S., McKeever TM, Churpek, M , Gonem, S and Shaw, D *Predicting outcome in acute respiratory admissions using patterns of National Early Warning Scores.* Clinical Medicine, Journal of the Royal College of Physicians of London, 2022. **22**(5): p. 409-15.

118. Forster, S., T.M. McKeever, and D. Shaw, *Effect of implementing the NEWS2 escalation protocol in a large acute NHS trust: a retrospective cohort analysis of mortality, workload and ability of early warning score to predict death within 24 hours.* BMJ Open, 2022. **12**(11): p. e064579.

119. Vardy, E.R., et al., *NEWS2 and the older person.* Clin Med (Lond), 2022. **22**(6): p. 522-524.

120. Dynesen, J., et al., *The 7-day mortality associated with an early warning score varies between age groups in a cohort of adult Danish emergency department patients.* Eur J Emerg Med, 2019.

121. Azraai, M., et al., *Observational cohort study of the triggers, diagnoses and outcomes of the medical emergency team (MET) response in adult psychiatry inpatients colocated with acute medical services in Australia.* BMJ Open, 2021. **11**(10): p. e046110.

122. Cioffi, J., *Recognition of patients who require emergency assistance: a descriptive study.* Heart Lung, 2000. **29**(4): p. 262-8.

123. Hughes, C., et al., *'Between the flags': implementing a rapid response system at scale.* BMJ Qual Saf, 2014. **23**(9): p. 714-7.

124. Ludikhuize, J., E. de Jonge, and A. Goossens, *Measuring adherence among nurses one year after training in applying the Modified Early Warning Score and Situation-Background-Assessment-Recommendation instruments.* Resuscitation, 2011. **82**(11): p. 1428-33.

125. Redfern, O.C., et al., *The association between nurse staffing levels and the timeliness of vital signs monitoring: a retrospective observational study in the UK.* BMJ Open, 2019. **9**(9): p. e032157.

126. Romero-Brufau, S., et al., *The fifth vital sign? Nurse worry predicts inpatient deterioration within 24 hours.* JAMIA Open, 2019. **2**(4): p. 465-470.

127. Douw, G., et al., *Capturing early signs of deterioration: the dutch-early-nurse-worry-indicator-score and its value in the Rapid Response System.* J Clin Nurs, 2017. **26**(17-18): p. 2605-2613.

128. Hoffman, R.R., B. Crandall, and N. Shadbolt, *Use of the Critical Decision Method to Elicit Expert Knowledge: A Case Study in the Methodology of Cognitive Task Analysis.* Human Factors, 1998. **40**(2): p. 254-276.

129. Gazarian, P.K., *Use of the critical decision method in nursing research: an integrative review.* ANS Adv Nurs Sci, 2013. **36**(2): p. 106-17.

130. Carayon, P., et al., *Work system design for patient safety: the SEIPS model.* Qual Saf Health Care, 2006. **15 Suppl 1**: p. i50-8.

131. Holden, R.J. and P. Carayon, *SEIPS 101 and seven simple SEIPS tools.* BMJ Qual Saf, 2021. **30**(11): p. 901-910.

132. Crooks, C.J., et al., *Pulse oximeter measurements vary across ethnic groups: an observational study in patients with COVID-19.* Eur Respir J, 2022. **59**(4).

133. Middleton, S., et al., *Factors affecting -individual task prioritisation in a workplace setting.* Future Healthc J, 2018. **5**(2): p. 138-142.

134. Brown, M., et al., *A survey-based cross-sectional study of doctors' expectations and experiences of non-technical skills for Out of Hours work.* BMJ Open, 2015. **5**(2): p. e006102.

135. Gonem, S., et al., *Dynamic early warning scores for predicting clinical deterioration in patients with respiratory disease.* Respir Res, 2022. **23**(1): p. 203.

136. Murray, A., et al., *Trajectories of the averaged abbreviated Vitalpac early warning score (AbEWS) and clinical course of 44,531 consecutive admissions hospitalized for acute medical illness.* Resuscitation, 2014. **85**(4): p. 544-8.

137. Kellett, J., et al., *Trends in weighted vital signs and the clinical course of 44,531 acutely ill medical patients while in hospital.* Acute Med, 2015. **14**(1): p. 3-9.

138. Odell, M., K. Gerber, and M. Gager, *Call 4 Concern: patient and relative activated critical care outreach.* Br J Nurs, 2010. **19**(22): p. 1390-5.

139. Ratcliffe, M., *Implementing the Call 4 Concern (C4C) service: patient and relative activated Critical Care Outreach.* 2023, Royal College of Nursing.

Appendix

1 Study Protocol for Quantitative Database Analysis

Chapters two to five of this thesis are dedicated to the three studies based on analysis of a large database of outcomes and demographics linked vital signs observations. The protocol approved by the HRA, University of Nottingham and Nottingham University Hospitals NHS Trust Information Governance and Research and Innovation bodies is detailed below.

Nottingham University Hospital Pseudo-anonymised Digital Health **Records Database**

Final Version 1.0 28/08/2019

Short title:	Investigating patient deterioration through clinical database analysis
IRAS Project ID:	*270837*
Database Sponsor:	University of Nottingham
Sponsor reference:	*19074*
Funding Source:	*Nottingham Hospitals Charity- William Colacicchi Fellowship*

1.1.1 Database Personnel And Contact Details

Sponsor:	University of Nottingham Ms Angela Shone Research and Innovation University of Nottingham East Atrium Jubilee Conference Centre Triumph Road Nottingham NG8 1DH
Chief investigator:	Professor Dominick Shaw Room 21 Clinical Sciences Building City Hospital Campus Hucknall Road Nottingham NG51PB Phone: 0115 823 1719 Email: dominic.shaw@nottingham.ac.uk
Co-investigators:	Dr Sarah Forster- studying for PhD in Respiratory medicine Sarah.forster@nottingham.ac.uk
Database Statistician:	Professor Tricia Mckeever Phone:01158231721 Email:Tricia.mckeever@nottingham.ac.uk

Database Coordinating Centre: Respiratory Research Group, School of Medicine

1.1.2 Synopsis

Title	Improving understanding of patient deterioration through statistical analysis of previously collected, pseudo-anonymised clinical data linked to outcomes to profile differences between patients who go on to experience a serious adverse event versus those who don't.
Short title	Investigating patient deterioration through clinical database analysis
Chief Investigator	Professor Dominick Shaw

Objectives	1- To profile patients who meet the outcomes of mortality and unplanned admission to ICU using vitals signs and additional routinely collected clinical data including, but not limited to, comorbidity score, admission diagnosis and pattern of previous admissions. 2- To determine whether pattern of changes in vital signs can be more predictive of serious adverse event 3- To determine the number of escalations and workload generated by the National Early Warning Score 2 across different specialty populations
Database Configuration	Single Centre
Setting	Secondary Care
Sample size estimate	All admissions between April 2015- June 2020- Approximately 10 million observation sets
Number of participants	All admissions between April 2015- June 2020- Approximately 500 000 patients
Eligibility criteria	Inclusion criteria- All admissions between April 2015- June 2020 aged 18 or older with one or more complete set of vital signs recorded following admission. Exclusion criteria- Patients aged under 18 in adult care Patients receiving end of life care
Duration of database	Database will be extracted from Nottingham University Hospitals clinical data between October 2019 and June 2020 and will cover all admissions between April 2015- June 2020. Data will be held from extraction until 7 years after completion of project as per University of Nottingham guidelines- total 10 years.

1.1.3 Abbreviations

CI	Chief Investigator overall
CRF	Case Report Form
GCP	Good Clinical Practice
NHS	National Health Service
P/GIS	Parent / Guardian Information Sheet
PI	Principal Investigator at a local centre
PIS	Participant Information Sheet
REC	Research Ethics Committee
R&D	Research and Development department
UoN	University of Nottingham
NEWS2	National Early Warning Score version 2

1.1.4 Table Of Contents

1.1.5 Database Background Information And Rationale

In hospital mortality across the NHS in the period April 2017 to March 2018 was recorded at 3.3% by the Office of National Statistics. This increased to 19.7% for patients with an unplanned admission to ICU [1] and 78% for patients who suffered an in-hospital cardiac arrest [2]. In addition, there is evidence to suggest that other patients in the same areas as patients who experience these serious adverse outcomes may themselves be at higher risk of deterioration due to diversion of resources [3]. This highlights the need to prevent these outcomes where possible by intervening at a lower level of clinical acuity. In order to do this it is necessary to have some form of monitoring system to provide an opportunity to detect at risk patients and to act. It was demonstrated in several studies in the 1980s/90s[4-6] that the majority of patients going on to experience a serious adverse outcome exhibited derangement of vital signs in the 8-12 hours prior to overt deterioration. This creates a potential window in which intervention may theoretically improve outcome if an appropriate response is triggered. This concept is referred to as recognise and rescue and can be thought of as a feedback loop with an afferent and efferent limb.

Over the last 20 years, efforts to develop and refine the afferent limb has instigated the development of early warning scores. Initially these were simple scores which triggered a response if a single vital sign was sufficiently deranged or if a nurse was concerned about the patient. These subsequently developed into more complex scores which are an aggregate of 6 core vital signs, with weighting linked to where each recorded vital sign observation sits in relation to a set normal range. The overall score is then applied to an agreed escalation protocol to trigger a clinical response. Although many scores have been developed, the release of the National Early Warning Score version 2 in December 2017 led to the standardisation of the afferent limb across the majority of hospitals in the NHS as mandated by NHS England.

Despite this, our understanding of NEWS2 in terms of real-world ability to predict patient deterioration before serious adverse event and make a difference to patient outcome remains limited [7, 8]. There have been a few smaller studies examining specialty cohorts, however the majority of studies used to justify its development and introduction are based on retrospective analysis of large datasets purely looking at the sensitivity and specificity in predicting outcome within a set period following the score in an unselected patient group.

Because early warning scores, and NEWS2 specifically, are now so embedded in monitoring systems in developed countries, removing them in order to carry out a prospective trial to assess impact would be both impractical and unethical. In addition, the only study to have incorporated a stepped wedge design to assess this suffered from being underpowered due to the rarity of the outcomes being measured within the hospitals in the study [9]. Therefore a different approach is needed in order to evaluate the impact of NEWS2 and guide future development to ensure the ongoing evolution of improved patient safety protocols.

Nottingham University Hospitals NHS Trust has the largest and most granular clinical dataset in the country. Over 2 million observations are recorded each year, with the ability to link to other information routinely collected throughout the inpatient stay in addition to workload generated and outcomes.

Previous studies have focussed on the area under receiver operating curve for an outcome within a set time of a set of vital signs observations [10-12]. Although this will tell you in a perfect system, with unlimited resources, and no other inputs, how effective a system would be in highlighting deterioration, due to the effects of limited clinical resources and alert fatigue it remains only part of the recognise and rescue picture.

For example, there are many patients with chronic alterations in physiological response due to underlying disease who may never go on to deteriorate but are persistently highlighted for review [13], thereby diverting resources from patients with potentially greater acuity or preventing nursing staff from carrying out other clinical work. There are also patients who are either highlighted by nursing staff due to clinical concern before they go on to have a high NEWS2 score or who deteriorate to have a cardiac arrest or unplanned admission to ICU without ever having abnormal vital signs. We also do not know what impact intervention has at a point where vital signs are raised. Is a NEWS2 of >4 too late to intervene in some patients? Or does that depend on the underlying characteristics of the patient involved? These are all important questions that the uniquely sophisticated data-set in Nottingham could potentially allow us to address creating an important contribution to the patient safety discussion going forward.

We therefore intend to examine the statistical performance of the latest National Early Warning Score in the Nottingham inpatient population. This will initially be an unselected cohort of all inpatients with more than one observation set following admission. A further analysis will examine the impact of chronic disease to determine whether there are specific changes in physiology which can be predicted in changing the vital signs observation patterns in these patients.

We will also analyse whether trends in vital sign variation can be utilised to improve the discriminative value of NEWS2 [14]. We will then investigate what other routinely collected digital healthcare information could be used to improve the score's performance, both in terms of providing additional information for clinical teams prioritising multiple high scores, and in the early escalation of a patient who is deteriorating but may not yet have deranged vital signs sufficient to trigger a clinical response.

The timing of this study also means we will be able to analyse the impact on workload and escalations of implementing NEWS 2 in a tertiary hospital that previously had a system tailored to the needs of each specialty within a fully electronic observations and task management system in order to determine what lessons can be learned going forward in the ongoing development of early warning systems. Evaluating the work-up detection ratio, or number of escalations[15] a clinician would need to respond to see an outcome, will allow an evaluation of systems impact as well as patient safety implications of NEWS2 through an interrupted time series analysis covering point of implementation.

1.1.6 Database Objectives And Purpose

The aim of this study is to develop greater understanding of patient deterioration through routinely collected clinical information. To meet this aim we will address the following objectives:

1- To profile patients who meet the outcomes of mortality and unplanned admission to ICU using vital signs and additional routinely collected clinical data including, but not limited to, comorbidity score, admission diagnosis and pattern of previous admissions.
2- To determine whether pattern of changes in NEWS2 can be more predictive of patient trajectory event than a stand-alone score
3- To determine the number of escalations and workload generated by the National Early Warning Score 2 across unselected and different specialty populations

1.1.7 Database Design

1.1.7.1 Database Configuration

Single centre. Vital signs observations data linked to background, in hospital progress and outcomes.

1.1.7.2 Database Management

Data will be linked and pseudo-anonymised by NHS data analysed before extraction. Extraction will occur between October 2019 and June 2020. No Patient identifiable data will leave the clinical system. It will then be saved in a secure folder on the R drive. Student and Academic Supervisors- CI and Statistician- will have access.

The Chief Investigator has overall responsibility for the database and shall oversee all database management.

The data custodian will be the Chief Investigator.

1.1.7.3 Duration Of The Database And Participant Involvement

Database Duration: 10 years

Participant Duration: As this is a retrospective data collection study there will be no involvement for participants and thus no participant duration.

1.1.7.4 End of the Database

Data will be kept for 7 years following the end of the project in line with university protocol. The end of the database will be 2029.

1.1.8 Selection And Withdrawal Of Participants

All admissions to Nottingham University Hospitals NHS Trust within the indicated period will be included if they meet the inclusion criteria. All patient identifiable data will be removed prior to extraction of data from clinical system.

No data will be collected specifically for this study. All data extracted is routinely collected in the course of the inpatient stay.

1.1.8.1 Eligibility criteria

No Patient Identifiable Details will be retained in the database therefore individual consent will not be sought in line with current protocols relating to use of routinely collected NHS data.

1.1.8.2 Inclusion criteria

Aged 18 or older in adult care
Admitted to NUH between April 2015-June 2020
With one or more complete set of vital signs recorded following admission

1.1.8.3 Exclusion criteria

Patients aged under 18 in adult care
Patients receiving end of life care

1.1.8.4 Expected duration of participant participation

There will be no duration of participation as this study looks at retrospective data only.

1.1.8.5 Participant Withdrawal

Not applicable

1.1.8.6 Informed consent

Not applicable- no patient identifiable data will be retained in the database following extraction from clinical system. This is in line with NHS information governance and clinical governance.

1.1.9 Data Collection Regimen

An NHS data analyst with access to the data involved as part of her role will access the clinical databases involved and link vital signs to clinical data including comorbidities, blood tests, diagnoses and outcomes before applying coding for individuals and visits. All data will have patient identifiers removed prior to extraction and will be coded to allow for linked analysis. The linkage for this code will be retained by the NHS data analyst extracting the data and will not be available to the research team. The linkage code will be kept in an encrypted file in a password protected folder of a secure NHS desktop in a locked office.

1.1.10 Access To Database

In line with previous permissions from the Nottingham University Hospitals Information Governance Team- Data will be kept in a secure folder on the University R drive and backed up

daily to an encrypted external hard drive that will be kept in a locked cupboard in a secure office.

1.2 Criteria for terminating the database

Due to the nature and use of the data that is due to be collected as part of this research database, there is no criterion for terminating the database.

1.2.1 Statistical Analyses

Analysis will be carried out by the PhD student (Sarah Forster) under the supervision of the statistician for this project Professor Tricia Mckeever.

We intend to examine the statistical performance of the latest National Early Warning Score in the Nottingham inpatient population. This will initially be an unselected cohort of all inpatients with more than one observation set following admission. A further analysis will examine the impact of chronic disease to determine whether there are specific changes in physiology which can be predicted in changing the vital signs observation patterns in these patients.

We will also analyse whether trends in vital sign variation can be utilised to improve the discriminative value of NEWS2 . We will then investigate what other routinely collected digital healthcare information could be used to improve the score's performance, both in terms of providing additional information for clinical teams prioritising multiple high scores, and in the early escalation of a patient who is deteriorating but may not yet have deranged vital signs sufficient to trigger a clinical response.

The timing of this study also means we will be able to analyse the impact on workload and escalations of implementing NEWS 2 in a tertiary hospital that previously had a system tailored to the needs of each specialty within a fully electronic observations and task management system in order to determine what lessons can be learned going forward in the ongoing development of early warning systems. Evaluating the work-up detection ratio, or number of escalations a clinician would need to respond to see an outcome, will allow an evaluation of systems impact as well as patient safety implications of NEWS2 through an interrupted time series analysis covering point of implementation.

1.3 Sample size and justification

The outcomes measured in this study are rare and this has caused issues with previous studies in this area being underpowered. Use of the maximum amount of data available will mitigate this possibility and allow analysis of sub-populations as well as the whole data-set.

1.3.1 Adverse Events

The occurrence of an adverse event as a result of participation within collection of data for this database is not expected and no adverse event data will be collected.

1.3.2 Ethical And Regulatory Aspects

1.3.3 Ethics Committee And Regulatory Approvals

The database will not be initiated before the protocol, consent forms and participant information sheets have received approval / favourable opinion from the Research Ethics Committee (REC), and the respective National Health Service (NHS) Research & Development (R&D) department(s). Should a protocol amendment be made that requires REC approval, the changes in the protocol will not be instituted until the amendment and revised informed consent forms and participant information sheets (if appropriate) have been reviewed and received approval / favourable opinion from the REC and R&D departments. A protocol amendment intended to eliminate an apparent immediate hazard to participants may be implemented immediately providing that the REC are notified as soon as possible and an approval is requested. Minor protocol amendments only for logistical or administrative changes may be implemented immediately; and the REC will be informed.

The database will be conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki, 1996; the principles of Good Clinical Practice, and the UK Department of Health Policy Framework for Health and Social Care, 2017.

The Health Regulations Authority have advised that REC opinion is not required for this study due to the fact that data being used has been previously collected as part of routine care and no additional data will be collected for the study.

As all patient identifiable data will be removed and not accessible by the research team informed consent is not required.

1.4 RECORDS

1.4.1 Case Report Forms

Not applicable to this study- No patient identifiable data will be accessed at any point.

1.4.2 Source documents

Not applicable to this study- we will not know the identity of patients

1.4.3 Direct access to source data / documents

Not applicable

1.4.4 Data Protection

All database staff and investigators will endeavour to protect the rights of the database's participants to privacy and informed consent, and will adhere to the Data Protection Act, 2018.. Access to the information will be limited to the database staff and investigators and any relevant regulatory authorities (see above). Computer held data including the database will be held securely and password protected. All data will be stored on a secure dedicated web server. Access will be restricted by user identifiers and passwords (encrypted using a one way encryption method).

All data will be stored on secure University servers. A front end will be created to link to secure data files and encrypted to 128bit AES.

Access to data file locations on university servers is granted only by written request from a senior manager and limited to research staff and PhD students where required. Access to secure data file locations is managed by University of Nottingham IT services and accessed by university username and password.

The database manager will be aware of their professional duty to maintain participant confidentiality at all times. No patient identifiable data will be stored

All data stored on University Servers is backed up by University of Nottingham I.T. Services on a daily basis.

A local back up is also carried out on a daily basis, encrypted using AES256 bit encryption and backed up to an external hard drive located and locked in a secure room. No patient identifiable data will be stored.

Electronic data will be backed up every 24 hours to both local and remote media in encrypted format.

1.4.5 Insurance And Indemnity

Insurance and indemnity for interventions conducted on participants to collect data for the database is covered within the NHS Indemnity Arrangements for clinical negligence claims in the NHS, issued under cover of HSG (96)48.

The University of Nottingham as research Sponsor indemnifies its staff, research participants and research protocols with both public liability insurance and clinical trials insurance. These policies include provision for indemnity in the event of a successful litigious claim for proven non-negligent harm.

1.4.6 Database Conduct

Procedures for the collection and storage of data for the database may be subject to systems audit for inclusion of essential documents; permissions to conduct the database; CVs of database staff and training received; local document control procedures; consent

procedures and recruitment logs; adherence to procedures defined in the protocol (e.g. inclusion / exclusion criteria, timeliness of visits); accountability of database materials and equipment calibration logs.

1.4.7 Database Data

Monitoring of database data shall include; data storage and data transfer procedures; local quality control checks and procedures, back-up and disaster recovery of any local databases and validation of data manipulation.

The Database Manager, or where required, a nominated designee of the Sponsor, shall carry out monitoring of database data as an ongoing activity.

1.4.8 Record Retention And Archiving

In compliance with the ICH/GCP guidelines, regulations and in accordance with the University of Nottingham Code of Research Conduct and Research Ethics, the Chief or local Principal Investigator will maintain all records and documents regarding the conduct of the database. These will be retained for at least 7 years or for longer if required. If the responsible investigator is no longer able to maintain the database records, a second person will be nominated to take over this responsibility.

The database documents held by the Chief Investigator on behalf of the Sponsor shall be finally archived at secure archive facilities at the University of Nottingham. This archive shall include all database records.

1.4.9 Discontinuation Of The Database By The Sponsor

The Sponsor reserves the right to discontinue this database at any time for failure to meet expected enrolment goals, for safety or any other administrative reasons. The Sponsor shall take advice as appropriate in making this decision.

1.4.10 Statement Of Confidentiality

No patient identifiable data will be extracted or stored.

1.4.11 Publication And Dissemination Policy

The data from the study is owned by the NHS. As this is a student project, a final study report will be available in the form of PhD thesis upon completion.

All publications will acknowledge the Nottingham Hospitals Charity as funder.

It is envisaged that the majority of outputs will be submitted for dissemination either in peer review journals or as conference abstracts prior to the completion of the PhD at the end of 3 years (July 2022).

Study data will be accessible by the chief investigator and supervisors only for the period of the PhD. Any future access to the study data would need to be requested as stated above.

1.4.12 User And Public Involvement

Stakeholders and patient have been involved in developing the aims and objectives for this study. Knowledge gained from this study will be disseminated to both stakeholders and patients through patient safety fora and conferences. Findings will be discussed within patient focus groups with the support of the Academic Health Sciences Network to determine further objectives based on information gained

1.4.13 Database Finances

1.4.13.1 Funding source

Nottingham Hospitals Charity

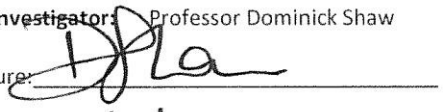
1.4.13.2 Participant stipends and payments

Not applicable- all data will be extracted from information collected during routine care in a hospital inpatient stay.

SIGNATURE PAGES

Signatories to Protocol:

Chief Investigator: Professor Dominick Shaw

Signature: 

Date: 29/8/19

1.4.14 References for study protocol

1. ICNARC, *Case Mix Programme Summary Statistics 2017/18*. 2018.
2. National Cardiac Arrest, A., *Key Summary Statistics 2017-18*. 2018.
3. Volchenboum, S.L., et al., *Association Between In-Hospital Critical Illness Events and Outcomes in Patients on the Same Ward.* JAMA, 2016. **316**(24): p. 2674-2675.
4. Schein, R.M., et al., *Clinical antecedents to in-hospital cardiopulmonary arrest.* Chest, 1990.

98(6): p. 1388-92.

5. Goldhill, D.R., S.A. White, and A. Sumner, *Physiological values and procedures in the 24 h before ICU admission from the ward.* Anaesthesia, 1999. **54**(6): p. 529-34.
6. Hillman, K.M., et al., *Antecedents to hospital deaths.* Intern Med J, 2001. **31**(6): p. 343-8.
7. Alam, N., et al., *The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review.* Resuscitation, 2014. **85**(5): p. 587-94.
8. McNeill, G. and D. Bryden, *Do either early warning systems or emergency response teams improve hospital patient survival? A systematic review.* Resuscitation, 2013. **84**(12): p. 1652- 67.
9. Haegdorens, F., et al., *The introduction of a rapid response system in acute hospitals: A pragmatic stepped wedge cluster randomised controlled trial.* Resuscitation, 2018. **129**: p. 127- 134.
10. Jarvis, S., et al., *Aggregate National Early Warning Score (NEWS) values are more important than high scores for a single vital signs parameter for discriminating the risk of adverse outcomes.* Resuscitation, 2015. **87**: p. 75-80.
11. Smith, G.B., et al., *The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death.* Resuscitation, 2013. **84**(4): p. 465-70.
12. Green, M., et al., *Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients.* Resuscitation, 2018. **123**: p. 86-91.
13. Forster, S., et al., *Investigating the discriminative value of Early Warning Scores in patients with respiratory disease using a retrospective cohort analysis of admissions to Nottingham University Hospitals Trust over a 2-year period.* BMJ Open, 2018. **8**(7): p. e020269.
14. Churpek, M.M., R. Adhikari, and D.P. Edelson, *The value of vital sign trends for detecting clinical deterioration on the wards.* Resuscitation, 2016. **102**: p. 1-5.
15. Kipnis, P., et al., *Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU.* J Biomed Inform, 2016. **64**: p. 10-19.
16. Sanchez-Pinto, L.N., et al., *Comparison of variable selection methods for clinical predictive modeling.* Int J Med Inform, 2018. **116**: p. 10-17.

1.5 Data Studies Approvals

The approvals for all three of the quantitative studies are covered by the same Health Regulations Authority and Nottingham University Hospitals approvals based on the study protocol described above.

Approval was given by the UK Health Research Authority (IRAS ID 270837) and Nottingham University Hospitals Trust's Caldicott guardian, Research and Innovation team and Information Governance department (Ref: DG20-000049-D and IG0025). As the study did not involve human participants and was

limited to routinely collected data anonymised prior to extraction, the HRA did not require research ethics committee review.

Full details of the regulatory approvals process can be found in a dedicated chapter containing a timeline of the approvals process and the Data Protection Impact Assessment and Data Sharing Agreements.

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Professor Dominick Shaw
B21 Clinical Sciences Building
University of Nottingham, City Hospital Campus
Hucknall Road
NG51PB

Email: hra.approval@nhs.net
HCRW.approvals@wales.nhs.uk

18 September 2019

Dear Professor Shaw

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Improving understanding of patient deterioration through statistical analysis of previously collected, pseudo-anonymised clinical data linked to outcomes to profile differences between patients who go on to experience a serious adverse event versus those who don't.**
IRAS project ID:	**270837**
Protocol number:	**19074**
Sponsor	**University of Nottingham**

I am pleased to confirm that HRA and Health and Care Research Wales (HCRW) Approval has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report

(including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The *"After HRA Approval – guidance for sponsors and investigators"* document on the HRA website gives detailed guidance on reporting expectations for studies with HRA and HCRW Approval, including:

- Registration of Research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **270837**. Please quote this on all correspondence.

Yours sincerely,

Michael Pate
Approvals specialist

Email: hra.approval@nhs.net

Copy to: *Ms Angela Shone – Sponsor contact.*

2 Regulatory Approvals

2.1 Timeline

Regulatory approval proved to be the most arduous task involved in this PhD. This project had been developed and refined over a period of 4 years and focused, as per the thesis on the analysis of previously collected clinical data, anonymised by an NHS data analyst prior to extraction, to develop our understanding of how early warning scores derived from vital signs are used to promote patient safety and how this data can be further developed in conjunction with other routinely collected clinical data to improve our discrimination of patients at risk of deterioration, to provide the potential to intervene and to reduce unnecessary intervention and workload.

Following submission of IRAS and supporting documentation to the HRA, approval was issued on 18th September 2019. This was in conjunction with support from the University of Nottingham as Sponsor, and NUH R&I. With Nottingham Hospitals Charity as primary funder for my fellowship.

I had previously been in contact with the Information Governance Lead at Nottingham University Hospitals, Rory King, in July 2019 for advice regarding information governance for this project and following receipt of approval from the HRA re-contacted him. He responded 2 weeks later with a request for further information, the new template for DPIA and a list of data points he would like us to consider restricting.

18th October- DPIA returned by Rory with positive comments and suggestions for improvement. The main question was regarding the data security at the University.

22nd October- After discussion with the digital research team at the university (Jasper Donelan) I sent the amended DPIA and the following suggestions for storage:

Teams: Encrypted, cloud-based service. No data leaves EU-jurisdiction (Dublin primary storage, back-up in the Netherlands).

Central Performance Storage: Local, University of Nottingham server. Encrypted and backed-up in two locations on campus.

1st November- Reply from Rory requesting cloud security questionnaire and data processing agreement. Cloud questionnaire sent to digital research team.

Request for information about cloud security followed up with digital research team on 6th, 12th and 15th of November.

20th November- received completed cloud questionnaire and forwarded to Rory as well as query regarding who needs to sign the data processing agreement.

25th November- follow up email to Rory met with out of office. 2nd December further email to Rory before 5th December email to information governance team.

6th December email from Martin Bakalarczyk explaining he would be looking at the project as Rory is unavailable. I therefore forwarded all documentation to Martin and arranged to meet and discuss.

13th December- Met with Martin and discussed project. Explained the clinical background in patient safety that is at the heart of the project and my involvement in several other projects at NUH including the introduction of NEWS2, the Cardiac Arrest Antecedent Audit and the current health foundation project looking at use of data to predict discharges. Discussed ways in which we could further anonymise data.

We agreed that I would alter the DPIA in line with our discussion and Martin would seek guidance on what level of ICD10 coding would be acceptable to anonymise patients sufficiently in the event of a jigsaw hack, which appears to be the main concern. However, Martin expressed doubts that any patient data should leave NHS servers. To cite precedent I pointed out that linked HES, CPRD and clinical trials data were all held in encrypted university storage in the same manner we were proposing having demonstrated the security of these systems.

16th December- Email to Martin returning the DPIA with alterations as discussed with him.

3rd January 2020- follow up email to check progress, Prof. Shaw copied in as becoming concerned regarding the impact of delay in accessing data on the project timeline. We received a reply on 6th January informing us he was out of office and to bear with him. Prof. Shaw replied to this email with a request to be contacted on his return from leave.

March 2020-July 2020- COVID 19 Clinical work- Approved break from research to return to work on adult intensive care at NUH.

End of September 2020- DPIA approved and signed off

End of November 2020- Alan Lowe informed me that the DPIA had been approved and signed off

December 2020- approached data warehousing team regarding extraction of data

January 2021- further discussions with data warehousing team regarding extraction of data- but following this unable to get much in the way of communication for several months.

June 2021- Database extraction completed and transferred by hand on encrypted CD which is kept in a locked cupboard in Professor Shaw's office.

2.2 DPIA

DATA PROTECTION IMPACT ASSESSMENT (DPIA) TEMPLATE –VERSION 2.0 (AUGUST 2019)

Introduction

A Data Protection Impact Assessment (DPIA) is a legal requirement for certain kinds of projects and activities. These include:

- ✓ new and innovative technologies such as artificial intelligence;
- ✓ where the data being used is sensitive and relates to an individual's health
- ✓ any systematic and extensive profiling
- ✓ monitoring of public areas
- ✓ where decisions are made to deny or allow access to services
- ✓ matching and combing data from more than one source
- ✓ where the data subject is considered vulnerable i.e. children, mental health
- ✓ where a data breach would result in significant harm to an individual or minor harm to many individuals
- ✓ automated decision making with significant effect upon individuals

The aim of a DPIA is to properly identify and assess the risk of an activity or project and the impact it may have upon an individual. It helps the Trust act in a lawful and risk-minimising way.

A DPIA is a dynamic process that should be revisited as systems/processes develop.

N.B. Data subject refers to the individuals whose data you intend to use.

If in doubt, complete a DPIA.

Project/System/Process Name	Recognition of the deteriorating patient

Key Personnel

	NUH Business/Operational lead	ICT lead (if applicable)	Supplier lead (if applicable)	Relevant Information Asset Owner/Assistant
Name	Professor Dominick Shaw	Gemma Housley		
Role	PI	Data analyst		
Email	Dominic.shaw@nuh.nhs.uk	Gemma.housley@nuh.nhs.uk		
Phone	0115 823 1719			

PLEASE COMPLETE EACH QUESTION FULLY. IF YOU THINK A QUESTION DOES NOT APPLY TO YOUR PROJECT/ ACTIVITY, EXPLAIN WHY.

Step 1: Aim of project/activity being undertaken

1. What is it that is being planned?

Provide a brief description and/or embed an Additional Service Request, Project Initiation Document or proposal.	Please see Protocol.  Investigating_patien t_deterioration_protocol

✓ **What is the nature of your relationship with the data subject whose data will be used?**

For example, do you provide direct care to the data subjects, are they your patients?	Dr Sarah Forster and Professor Dominick Shaw provide direct care to patients at NUH and therefore have access to data in clinical role. Gemma Housley is a senior data analyst jointly employed by NUH and the AHSN Professor Tricia Mckeever is a statistician involved in several clinical trials currently ongoing at NUH. Employed by the University of Nottingham.

✓ **Why are we doing it?**

Summarise why there is a need for implementation or change and the benefits it will realise.	We aim to improve understanding of patient deterioration through statistical analysis of previously collected, pseudo-anonymised clinical data linked to outcomes to profile differences between patients who go on to experience a serious adverse event versus those who don't. Using this information, protocols for identifying deterioration can be improved, potentially detecting those who are currently missed, and resources more efficiently channelled to patients at risk of deterioration. Previous analysis by this team on a smaller scale has already fed back into service improvement decisions at NUH. Any findings from this project would

2

	also be directed back into the national discussion regarding the ongoing development of NEWS. Specific questions include: ✓ Number of escalations and workload generated by Nottingham EWS and NEWS2 by specialty; including in hours versus out of hours escalations- This will allow us to understand when and where workload is being generated and if there is a higher risk of deterioration. ✓ In medical patients- profile of patients who die versus those who don't- this will allow us to understand our patient population in terms of prioritising review with the intention of feeding any findings that may improve discrimination of NEWS2 back into ongoing NEWS development. ✓ Across all patients- Profile of patients who have a serious adverse event (ICU/ Death) without a high EWS- to see if there are factors which could be used to predict. From the recently completed cardiac arrest antecedent audit and published literature surrounding EWS this is a small but significant group that if we could find markers/ risk factors to identify these patients it could be fed back into the ongoing development of NEWS ✓ Across all patients- Can change in observations over time be used to predict outcome more effectively? This has the potential to both improve detection of patients who deteriorate without significant derangement from normal, but who have significant changes from what is normal for them. It also has the potential for screening out patients who have baseline derangement due to chronic conditions thereby reducing unnecessary observations intervention. Again any findings would be fed back into local patient safety processes the national discussion regarding ongoing development of NEWS in order to maximise patient management and safety.

4. Individuals need to be told how their information is processed.

Is this covered by an existing fair processing information or leaflet? If Yes,	https://www.nuh.nhs.uk/gdpr

3

provide details. If No, go to Q5	Trust website information for patients: ***Research*** We may also ask you to volunteer to take part in health research and, if you do want to take part, we will ask for your agreement to use your data for this research. **We may also use your data, or part of it, for other reasons:** Receive funding and keep track of spending Teach and train our staff **Develop and improve care for patients in the future through research** Manage and plan our services

5. Have they been informed and consented to this use?

a. Have you consulted the data subject or their representative about using this data? If not, please explain why you haven't consulted them?	N aiming to keep o-patients anonymised to study team. Data all collected as part of routine care.
b. Please provide details and an example of how this consent was given?	Precedent from other studies within the NHS using vital signs observation data have proceeded without obtaining individual consent following discussion with HRA.
c. Explain why you believe they would consider the proposed new use of their data as being reasonable or expected?	We are using the data to learn from what has happened in the clinical care and trajectory of patients in order to understand our systems better so we can feed this into improved design and implementation going forward. Any findings will be fed back into patient safety and management decision making to further develop and improve patient care.
d. How will you tell patients/staff how their data is being used and if not, why not?	Not informing directly- as we are anonymising to study team (pseudo- anonymising in practice) and all data is routinely collected as part of their normal inpatient stay. Patients are informed on the NUH website that their data may be used in this way. https://www.nuh.nhs.uk/gdpr

4

✓ **Has an assessment been made that the information collected is the minimum required to meet the aim of the project?**

a. Use of data should not be the first resort if the objective can be achieved without its use. You must justify why the use of all the data is necessary and proportionate. For example, do you need to use all the fields, can you not achieve the same objective with fewer data fields and/or a smaller data set?	Regarding sample size- because outcomes are so rare- previous analysis of smaller data sets have been underpowered- particularly when looking at surgical patients as we would be to answer questions 1, 3 and 4. To make the best use of the data in generating data and findings that can be applicable in other settings, we need the large size of database, the spectrum of specialties and the seasonal trends that 5 years provides. The fields specified for use were chosen following from previous analysis of NUH data as part of service improvement in 2013. They provide the necessary context for interpreting analysis. Without this we could not interpret outcomes.
b. Has consideration been given to how the same objective or outcome may be achieved without using this data or using less data or employing a different method - explain in full?	A full review of the literature has been performed and we have been developing the project carefully over the last 4 years, including discussions with specialists in this field to discuss practical statistical approach and data needed to perform the analysis. We can only identify how trends in the data relate to rare but clinically important outcomes such as admission to ICU and death through having sufficient breadth of data to account for confounders in how vital signs observations relate to deterioration and sufficient numbers of observations to power the study. This project is only possible because of the granularity of data in the Nottingham system. Note- HRA approval has been obtained for this project

✓ **Is this processing using novel technology or for a novel purpose that would be of public interest or attract criticism. Explain your reasons?**

a. Is the technology or activity new to the Trust or is it a recent development?	Observational- no new technology
b. Where else in the technology used, have they completed a DPIA that may be useful for background information, if	We have not previously completed a DPIA

so please provide?	

Step 2: Data description and use

8. What type of data is being processed?

Fully describe ALL the data that will be used, for example what data fields are being used e.g. NHS No, date of birth, name etc.?	**Proposed data fields:** Pseudo-patient ID (to capture patients with multiple admissions) Pseudo-visit ID Age-group (5 year band) Previously used 18-20, 21-25, 26-30 etc Gender Admission date Admission ward Location at time of observation Specialty Next ward Outcome (in hospital death or survived to discharge) Time since first observation of death or discharge Discharge destination Length of stay Comorbidity Score (Charlson index) Resuscitation status (if recorded- from date when nervecentre started recording this) Chart type (admission/ adult/ end of life/ renal/ resp etc) Frequency of observations (60/120/180 minutes etc) Overdue flag Overdue time Hour of day of observation Day of week of observation (including whether bank holiday or not) Minutes since first observation Respiratory rate Respiratory rate score Oxygen saturations

	(Oxygen saturations score after June 25th 2019 switch over) Inspired Oxygen Inspired Oxygen Score Temperature Temperature Score Blood Pressure Blood Pressure Score Heart Rate Heart Rate Score AVPU AVPU Score Urine Outuput (Urine output score- before June 25th 2019) EWS (NEWS2 after 25th June 2019) Escalated to Reason not escalated/escalated if any) Admission Primary diagnosis Admission primary ICD10 Dominant primary diagnosis Dominant primary ICD10 Discharge Primary diagnosis Discharge primaryIC
	Sepsis flag Time from first observation made medically fit for discharge (time to medically fit LOS is relevant) (frailty flag if possible) Most recent blood gas CO2- evaluate scale 2 of NEWS2 Most recent blood gas HCO3- evaluate scale

9. Does this include any of the following special categories of personal data?

Data concerning health	☒
Data concerning sex life or sexual orientation	☐

DATA PROTECTION IMPACT ASSESSMENT (DPIA)–Recognition of the deteriorating patient November 2019

Genetic or biometric data	☐
Racial or ethnic origin	☐
Religious or philosophical beliefs	☐

10. Approximately how many individuals will be in the dataset?

<100	☐	100-500	☐	500-1,000	☐	1,000-10,000	☐	10,000-50,000	☐	>50,000	☒

✓ **What volume of data will be involved?**

How large and expansive are the records sets being used, what will it consist of? What geographical area will the data be drawn from or cover?	Admissions to Nottingham University Hospitals NHS trust between April 2015 and June 2020. Estimated to include approximately 500,000 patients, 10 million observations.

✓ **What is the source of this data?**

a. If the data is being taken from an existing system, identify what system that is and what was the originally purpose that data was collected for? How will this data be accessed	Data to be collected from the clinical database at Nottingham University Hospitals NHS trust. It was originally collected as part of routine patient care and stored on Medway, Nervecentre or Notis. It will be accessed by a data analyst who works for the trust and would normally have access as part of her job role for the NHS.
b. If it new data that is being collected, describe how this data collection will be done i.e. electronic form, paper form etc.?	No new data will be collected

✓ **How will this data be used?**

a. Will this data be used or combined with other data sets, if so what are these other data sets?	Only data from Nottingham University Hospitals NHS trust will be used.
b. What will this data show you that is relevant to the project aim and purpose	The data is the foundation of the project. The data will answer the questions as stated in question 3 above

8

DATA PROTECTION IMPACT ASSESSMENT (DPIA)–Recognition of the deteriorating patient November 2019

14. Duration of processing

What is the duration of this processing? Is this one off processing or will it continue for a specified period?	Data processing to continue until May 2022. This is a period of research for Dr Sarah Forster, an ICU and Respiratory trainee employed by the trust to carry out this research funded by Nottingham Hospitals charity as part of a PhD.

15. How long will the data be kept and how will it be deleted?

✓ NHS data needs to be retained in accordance with the NHS Records Management Code of Practice. Has provision been made to ensure you are able to accommodate this? If No, describe how the data will be stored.	The pseudo-anonymised data without patient identifiable data will be stored in an encrypted folder on a secure university research server. It will remain there throughout the study. A backup will be made on a daily basis onto an encrypted external hard-drive which will be placed in a locked drawer in a locked office on the City Hospital Campus site of the University of Nottingham when not in use. In compliance with the ICH/GCP guidelines, regulations and in accordance with the University of Nottingham Code of Research Conduct and Research Ethics, the Chief or local
✓ If data is being processed by a third party, how will we ensure data is deleted when required?	**No data will be accessed or processed by a third party**. Data will be accessed and processed by Gemma Housley, and Sarah Forster (after it has been pseudo-anonymised). Statistics supervision will be provided by Professor Tricia Mckeever who is also statistician for several clinical trials in conjunction with NUH and is a professor of Medical statistics and

16. How has the required data been minimised?

You are required to minimise the amount and level detail of any data set. For example, dates of birth should not be	Use of age-bands instead of age Avoiding collecting demographics not necessary for outcomes of this work. Discharge destination limited to supported versus unsupported

used where age would provide sufficient information to achieve the project aim. Explain how each field required is required to achieve the project aim and why it cannot be minimised further?	Following submission of the data fields described in question 8. Clarification was requested by Rory King (Information Governance) on the following: - Pseudo-patient ID (to capture patients with multiple admissions)- Allows us to determine whether admissions we are analysing are stand-alone or a continuation of the same acute problem. Also how the behaviour of people with multiple admissions changes and whether that happens in a predictable way within their NEWS2 scores. - Pseudo-visit ID- as above - Age-group (5 year band) Previously used 18-20, 21-25, 26-30 – Age is an acknowledged confounder in the questions we are asking and will inform regarding differences in behaviour- We are using age-bands rather than age to assist in anonymity - Gender- Again an important confounder - ***Admission Month*** and day of week - Admission ward type- (instead of specific ward to reduce potentially identifiable information) admissions area/ high dependency/ specialty ward - Location at time of observation- current ward - Specialty- gives information that ward doesn't- may have chronic conditions specific to specialty, also different specialties have different protocols despite the hospital-wide EWS and outliers have different risk factors. - Next ward type (specialty ward/ outlier flag/ high dependency area)- allows us to capture if a patient was sent to theatres for an operation or ICU following a specific pattern of observations- admission to ICU is also a primary outcome as part of the composite outcome of serious adverse event. - Outcome (in hospital death or survived to discharge)- Death in hospital is a primary outcome for the work we are planning to do (along with admission to ICU it is classed as a serious adverse event) - Time of death or discharge since first observation- looking at the predictive nature of the score and the factors surrounding its use is

and if so how? What steps have been taken to minimise the risk of re-identification of anonymised or pseudonymised data?	this. The linkage key will not be available to the study group and will be deleted once extraction is complete. Data needs to be pseudo-anonymised in case any patterns are identified which raise patient safety concerns regarding a particular group of patients. This would then allow the appropriate clinical team at NUH to use the data linkage key to identify and investigate. Any concerns would be reported to the Caldicott Guardian for review.

Step 3: Data Security

All personal data must be stored, transmitted and deleted securely.
NOTE: The supplier or third party may be able to advise on this section. Please provide details for all questions.

18. Where will these data be stored?

a. Will the data be stored on NUH servers or servers external to the Trust?	External
b. If external, where will it be stored, will this be the UK, EU or elsewhere?	UK
c. Will the storage be controlled by another party such as a product/ platform supplier?	Central Performance Storage of the University.
d. If the data storage or processing is being done by a third party, what certifications do they hold i.e. ISO 27001 or other certifying bodies? When were they, and the proposed storage mechanism, subjected to an external penetration test and is a report available?	**Await decision by email to clarify and attach certificates
e. If the data is stored on a cloud platform, please provide details, and complete and embed/attach a cloud	N/A

12

security questionnaire.	

19. How will this data be secured during storage and when being moved?

a. Will it be encrypted when stored and/or moved, if so what type of encryption will be employed?	Encrypted while stored. Encrypted on NUH memory stick when moved.
b. Will it be on a server protected by firewall and network intrusion detection?	Yes
c. What technical controls are in place to prevent hacking of the data by unauthorised persons?	Data on the R Drive resides on a SAN, to which access is controlled via specific a specific access control mechanism, so when you request space on the R drive then only the people you specify are allowed to view the data. You need to have a valid UoN username and password to be able to access the R Drive and must be in the correct AD group which in turn grants read or read/write permissions. Addition and removal from these AD groups can only be performed by administrators. Access to the R drive is restricted to authenticated users, both on campus via authenticating on a workstation or remotely via Remote Desktop into a machine on campus, via Citrix or via VPN, all of which require a valid UoN account and are soon to utilise Multi Factor Authentication. Data on the R Drive remains within the UoN campus and the back ups of the data also reside within the campus. Internally the Workstation are protected by a central patching mechanism via SCCM to ensure security updates are pushed out to machines and workstation and servers have Sophos Anti-virus installed to protect against malware and viruses. We have border firewalls in place with a default deny policy and a robust Firewall Change Request process to create rules to allow legitimate traffic access when needed. As part of this process the machines are security scanned and audited regularly. We have a third part to provide a 24/7 managed security service using security feeds based on both CTM sensors within the campus and agents

	on servers (including file servers)
d. When being moved will it be secured through encrypted file transfer, secure transmission through SLL/TLS/SHS, please explain the specific technical standards that will apply?	Transfer will be done by Gemma Housley- Using NUH encrypted memory stick, password protected Excel files saved from my NUH machine, transferred to UON machine directly onto the R drive and files deleted from the memory stick. Memory stick secured in locked drawer in locked office.

20. Who will have access to this data and how will this access be controlled?

a. Will the data be kept on a system that is password controlled, what is the password length and how often does it have to be changed? Who will administer these access controls?	Access to the R drive is restricted to people directly involved in the project. In this case those with access will be Professor Shaw (data custodian and CI), Professor Mckeever (Statistician), Gemma Housley (NHS/AHSN data analyst), Sarah Forster (Clinical Research Fellow, Respiratory Medicine, NUH)
b. What other security measures are in place, such as physical security, multiple factor authentication?	Study team will require their university authentication and file passwords to access data. Backup to encrypted hard-drive that will automatically wipe with 5 incorrect password attempts. Password known only to Sarah Forster. Encrypted hard drive kept in locked drawer of locked office.

21. If you are using devices such as laptops to access data, how are these secured and managed?

Are all laptops and PCs encrypted and protected and how this is done?	No laptops will be used. All PCs are password protected. Data will not be held directly on PC but accessed through further encryption on university central storage.

22. What will happen to the data at the end of the project/activity or end of contract with a third party? Will it be returned or deleted and how will this be done?

Most contracts specify what happens to data at the end of contract. If this is not subject to contract, how will you ensure	As the data is part of a PhD the university stipulates it be kept for 7 years following completion of the PhD. Data will be archived in the University

the data held by any third party is deleted? Embed extract of contract as necessary with highlighted sections.	Secure Research Archive under closed access terms.

Step 4: Data Use and Sharing

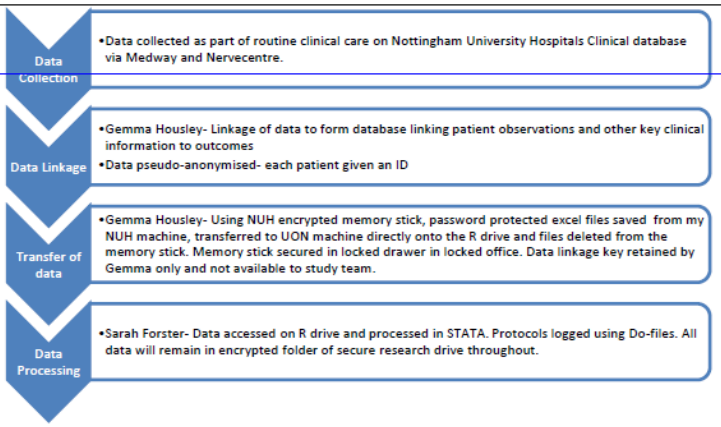
23. Will this data be shared with anyone else?

a. If yes, explain who these other parties are and why the data is being shared?	No
b. What is the statutory reason for this sharing?	Not applicable

24. Are other people processing this data?

a. If a third party such as a company is storing or otherwise managing or using our data, please explain what they doing and why they are doing it?	No third party involved in processing data. This will be done by Dr Sarah Forster, an employee of NUH for the duration of the project.
b. If we are using a third party product that requires maintenance where they access our networks, explain how this will be managed (will they remotely connect, how will this access be managed).	Not applicable

25. Describe the data flow?

a. Please embed or attach a simple flow map or visual description of how the data is collected, moved and used?	
b. Are there security or data protection concerns in any of the data flow stages you identify? If so, please indicate where and what steps you taking to reduce these risk?	Any concerns regarding transfer of data are being minimised by the use of an NUH encrypted memory stick to transfer from hospital to university storage. Memory stick will not leave secure office during process and will be immediately wiped on completion of transfer.

Step 5: Processing by or with a third party

a. If you are using a third party company or organisation to process, store or otherwise interact with this data, what is the arrangement between the Trust and the third party concerned?	Data will be stored on research 'R-drive' at University of Nottingham. This is part of the on-site, Central Performance Storage held at Kings Meadow Campus in Nottingham and used for sensitive research data. Used for all clinical trials data for studies conducted between the University and the NUH. See document...

	Central Performance Storage.pdf
b. What activities will the third party carry out i.e. storage, transport, processing of data on their platform?	University of Nottingham will be responsible for storage of data. Sarah Forster will be responsible for processing. Professor Shaw, an honorary consultant at NUH, will be Data custodian.
c. If this is carried out under a contractual agreement, please provide a copy of the contract and any other documents covering this arrangement.	Sponsor agreement and insurance attached from University of Nottingham
d. What steps or measures will you put in place to manage these risks? What measures will you take to ensure processors comply? PLEASE EMBED/ATTACH COPIES OF ANY CONTRACT/AGREEMENT.	Sponsorship agreement and insurance attached. HRA approval with documented agreement by University of Nottingham as part of approvals process.
e. If the data is being processed outside of the UK, please explain where?	N/A
f. If the data is processed outside of the EEA, what safeguards will be in place?	N/A

Step 6: Consultation

Consider how to consult with those who have an interest in this project/system:

a. Describe when and how you will seek individuals' views or justify why it's not appropriate to do so.	This is a study examining data that is collected routinely. Patient-id will be anonymised to the study team. Information regarding use of patient data available to all on NUH website. https://www.nuh.nhs.uk/gdpr
b. Who else do you need to involve within	Approvals in place from HRA, University of Nottingham Sponsors team,

your organisation?	Nottingham University Hospitals NHS Trust R&I. Also contacted Caldicott Guardian, Jeremy Lewis, who advised gaining HRA approval which we have done. 270837 - Letter of HRA Approval - 2019
c. Do you need to ask your data processors to assist?	Coordinating with data analysts employed by NUH.
d. Do you plan to consult information security experts, or any other experts?	No

Step 7: Assess necessity and proportionality

Describe compliance and proportionality measures, in particular:

a. What is your lawful basis for processing?	Yes	No
Consent	☐	☒
Processing is necessary for the performance of a contract	☐	☒
Processing is necessary for compliance with a legal obligation to which the controller is subject;	☐	☒
Processing is necessary in order to protect the vital interests of the data subject	☐	☒
Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested the controller	☒	☐
Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party NB. Not for public bodies.	☐	☒
b. Explain how the processing actually achieves your purpose?	By addressing the questions detailed in question 3 the project aims to contribute to the ongoing discussion around development of systems to recognise deteriorating patients in time to	

DATA PROTECTION IMPACT ASSESSMENT (DPIA)–Recognition of the deteriorating patient November 2019

	intervene. Good quality research in this area has the potential to influence the development of future iterations of NEWS. Now EWS is being mandated on a national level it is necessary to create the evidence base to drive this forward.
c. Is there another way to achieve the same outcome, give details of alternative you have rejected?	No- using a data set of this size is the only way to find statistically significant results with the outcomes in use. We need to use these outcomes as they are clinically relevant and consistent with current literature.
d. How will you prevent function creep, what are the governance arrangements around this activity/project?	We have a study protocol approved by the HRA, University of Nottingham and NUH R&I. Data will only be used in line with this study protocol.
e. How will you ensure high standards of data quality, please explain why all the data fields are necessary to achieve the objective?	Sarah Forster has completed training in advanced statistics and DOME at the University of Nottingham. She will be working under supervision of data custodian, Professor Shaw, and professor of Medical statistics and Epidemiology at the University of Nottingham Professor Mckeever (Professor Mckeever is analyst on several NUH clinical trials at present).
f. Please explain why a smaller amount of data cannot be used?	Outcomes are rare. Other studies in the literature with smaller amounts of data have been underpowered.
g. What information will you give individuals informing them of what you are doing with their data? How will you help to support their rights? PLEASE EXPLAIN YOUR THINKING.	Data is being analysed for the purposes of developing patient safety and improving patient care as per the information provided on the NUH website regarding use of patient data for research centred around patient safety.

Step 8: Identify, assess and mitigate risks

Risk ref.	Risk	Consequence(s)	Risk rating now			Mitigation/Recommended Actions	Risk rating after mitigation			Status
			Likelihood	Consequence	Risk score		Likelihood	Consequence	Risk score	
PR01	Data subjects are not aware of how their information is processed (may not be clear from the privacy notice that this type of processing may be undertaken)	Potential breach of GDPR Article 12 (transparent information).	3	4	12	The purpose is covered by the NUH privacy statement in broad terms. It is not practical to list every potential use in detail. Patient identifiable data is not processed outside of NUH employees or systems. Pseudo-anonymised data will be handled with high level of encryption on University computer with access by people named in this document only.	2	4	8	CLOSED
PR02	Public distrust about how information is used can damage an organisation's reputation and lead to loss of business.	Potential breach of GDPR Article 5 (lawful processing).	2	3	6	As above	1	3	3	CLOSED
PR03	Data security in transfer from NUH to Nottingham University systems	Potential data security breach	2	5	10	Use of encrypted memory stick for transfer of data. Use of encrypted folder on secure server on site at Kings Meadow Campus, Nottingham. Checks made on security level provided	1	5	5	CLOSED

DATA PROTECTION IMPACT ASSESSMENT (DPIA)–Recognition of the deteriorating patient November 2019

Risk ref.	Risk	Consequence(s)	Risk rating now			Mitigation/Recommended Actions	Risk rating after mitigation			Status
			Likelihood	Consequence	Risk score		Likelihood	Consequence	Risk score	
PR04	Identification of patients from information on database	Potential breach of data protection. Additional impact of feeding into PR02	2	4	8	Pseudo-anonymisation of data before transfer from NUH system. Limiting fields as detailed above. Only using data necessary to achieve research aims. Signed data processing agreement.	1	4	4	OPEN
PR05	Identification of pattern of clinical concern and unable to put patient safety measures in place	Unable to act on identified patient safety risk	2	4	8	Pseudo-anonymised ID with Gemma Housley retaining linkage key- to be used only after approval of ID to mitigate risk and allow appropriate clinician to address any systemic concerns identified.	1	4	4	CLOSED

Step 9: Sign off and record outcomes – IG TEAM TO ACTION

CALDICOTT GUARDIAN/SIRO OPINION

APPROVED	☒		
CONDITIONALLY APPROVED (PENDING THE COMPLETION OF THE FOLLOWING ACTIONS)	☐		
REJECTED (FOR THE FOLLOWING REASONS)	☐		
CALDICOTT GUARDIAN/SIRO NAME/SIGNATURE	Dr Jeremy Lewis	DATE	28/09/2020

DATA PROTECTION OFFICER (DPO)

DPO COMMENTS	Comments above for consideration and responses to be fed back to the Data Guardian (Alan Lowe) The Trust now has an agreement with UoN for the safe and secure transfer, storage and management of data to the TRE for research purposes. This along with directions for the use of data for research on Covid means this project can be approved.	**Notes** DPO needs to be satisfied that: - Risks have been identified and adequately mitigated - Any actions for conditional approved will be resolved - The ICO is consulted where high risk processing is not mitigated - DPIA is reviewed to ensure ongoing compliance. If DPO decision departs from SIRO/Caldicott Guardian, DPO must explain what has been done to resolve any discrepancy.
DPO NAME/SIGNATURE	Digitally signed by Alan Lowe Date: 2020.09.28 15:54:19 +01'00'	DATE 28/09/2020