PREDICTING CLINICAL DETERIORATION

A thesis submitted for the degree of

Doctor of Philosophy

In the Discipline of Acute Care Medicine
Adelaide Medical School
Faculty of Health and Medical Sciences
The University of Adelaide

James Malycha

August 2021

TABLE OF CONTENTS

Abstract…..……4

Declaration……...…….5

Acknowledgments…..……..…….7

Format of thesis……8

Thesis signatures……10

Introduction……11

Chapter 1. Evaluating variables for potential inclusion in Hospital Alerts Via Electronic Noticeboard (HAVEN)……...19

Variables associated with unplanned general adult ICU admission in hospitalised patients: protocol for a systematic review……..20

Patient centred variables with univariate associations with unplanned ICU admission: a systematic review…….…….34

Chapter 2. Evaluating the prognostic value of fractional inspired oxygen for potential inclusion in HAVEN…….…..98

The effect of fractional inspired oxygen concentration on early warning score performance: a database analysis…….99

Chapter 3. Evaluating HAVEN in the ward environment……...139

Testing a digital system that ranks the risk of unplanned intensive care unit admission in all ward patients: protocol for a prospective observational cohort study……..140

Evaluation of a digital system to predict unplanned admissions to the intensive care unit: a mixed-methods approach……...……155

Chapter 4. Validating HAVEN……...184

Detecting deteriorating patients in hospital: development and validation of a novel scoring system…….…185

Chapter 5. Towards better outcome measures for HAVEN…….……258

Establishing consensus definitions for the deteriorated ward patient – a protocol…...……259

Chapter 6. Automated quantification of the clinical workload associated with systems like HAVEN……..…274

Using real-time location devices (RTLD) to quantify off-unit adult intensive care registrar workload: a one-year tertiary NHS hospital prospective observational study…….……275

Future Directions……..……291

Conclusion……..…..……..295

Appendix A - Presentations at national and international meetings……...299

Appendix B - Prizes awarded during candidature……300

Appendix C - Grants awarded during candidature……..301

Appendix D - Other publications during candidature……..302

Abstract

This thesis describes the development of a prognostic algorithm that uses Electronic Patient Record (EPR) data to predict potentially avoidable adverse events (e.g., cardiac arrest/unanticipated Intensive Care Unit (ICU) admission) in sufficient time so that interventions can take place in patients admitted to the hospital ward. The system is called Hospital-wide Alerts Via Electronic Noticeboard (HAVEN). The thesis is composed of six chapters: evaluating variables for potential inclusion in HAVEN (chapter 1), evaluating the prognostic value of fractional inspired oxygen for potential inclusion in HAVEN (chapter 2), evaluating HAVEN in the ward environment (chapter 3), validating HAVEN (chapter 4), working towards improved outcome measures for HAVEN (chapter 5) and the automated quantification of the clinical workload associated with systems like HAVEN (chapter 6).

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

James Malycha
31 August 2021

Acknowledgments

Firstly, thanks to Professor Adam Deane, who facilitated applications for both my job in Oxford at the John Radcliffe Hospital (JRH) Intensive Care Unit (ICU) and this PhD. It was at the JRH that I worked under Professors Duncan Young and Peter Watkinson, both senior ICU consultants and co-directors of the Kadoorie Centre for Critical Care Research and Education, University of Oxford. Consequently, I had the opportunity to observe Peter during a non-clinical research week and was first introduced to the Hospital Alerting Via Electronic Noticeboard (HAVEN). Neither Peter nor Duncan took much initial interest in my desire to do research but were eventually persuaded to supervise my first project, which formed the beginning of this thesis. Thereafter I was fortunate to receive financial support via the HAVEN Project as a Clinical Research Fellow within the Kadoorie Centre. I am grateful to Peter and Duncan for giving me that opportunity and I am fortunate to have had Supervisors of such esteem. I look forward to ongoing collaborations with Peter whilst wishing Duncan a very happy retirement. Thanks to Professor Guy Ludbrook, my Principal Supervisor. Guy's academic guidance has been essential, especially in providing a link to Australian academia. We enjoyed several meals during his visits to the UK and I hope these will continue in Australia, along with our developing research collaborations. Thanks to Professor Sandra Peake, Director of the Queen Elizabeth Hospital ICU, who has ably mentored and facilitated my research career during the early part of my consultant life in Adelaide. Thanks also to Mr Lukah Dykes and Mr Adam Phillips for their support, hard work, and enthusiasm over the last year as we attempt to implement HAVEN into the South Australian hospital ward environment.

I was fortunate to work with excellent people in Oxford, both in research and clinically. I would like to thank Dr Tim Bonnici for his many hours of support during the early phase of my PhD. Dr Oliver Redfern arrived in Oxford in 2016 to work on HAVEN and became a trusted colleague and friend. My thanks to Oliver for his important and ongoing guidance. Thanks to all my ICU colleagues at the JRH but specifically Dr Jim Day, Dr Chris Andersen, Dr Elaine Armstrong, Dr Jodie Smythe, Dr Mirae Harford, Dr Stuart Benham, Dr Douglas Wilkinson and Dr Laura Vincent. It was through their hospitality, friendship, and generosity that my family were able to thrive in Oxford, ultimately enabling me to pursue this PhD.

Thanks to Peter and Jenny Malycha, who as loving parents have endured the many lows and occasional highs of my academic life. Their support and encouragement have been the foundation. Thanks to Donald and Nimfa Farrell, whose support during our time in Oxford was

essential. Their many visits were greatly appreciated and made it possible for us to stay in the UK for as long as we did. Thanks to my sister Sarah, her husband Joe and my nephew Will, for their comradeship during our five amazing years in the UK. Thanks to my wife Mary for her patience and support during this PhD but also during the long years of my specialist training. Things have recently become easier but Mary's ability to keep our family functioning during those earlier, more challenging times are here publicly acknowledged (and greatly appreciated). Finally, I would like to mention my children Teddie and Flora, who may one day glance at this work with appropriate disinterest. Both born in Oxford and much loved.

Format of thesis

This thesis is by publication, supplemented by narrative, as per University of Adelaide guidelines. The thesis has six chapters, each with an introduction, a manuscript and protocol where relevant. The thesis contains eight manuscripts: three protocols, one review and four original studies. At the time of submission, six had been published and two had been submitted for publication. The manuscripts are presented in the form they were published in UK English and were non solicited. References follow each manuscript.

The publications are as follows:

Malycha J, Bonnici T, Sebekova K, Petrinic T, Young JD, Watkinson PJ: **Variables associated with unplanned general adult ICU admission in hospitalised patients: Protocol for a systematic review**. *Systematic Reviews.* 2017; 6:4–7

Malycha J, Bonnici T, Clifton DA, Ludbrook G, Young JD, Watkinson PJ: **Patient centred variables with univariate associations with unplanned ICU admission: A systematic review**. *BMC Medical Informatics and Decision Making.* 2019; 19:4–7

Malycha J*, Farajidavar N*, Pimentel M, Redfern O, Tarassenko L, Meredith P, Prytherch D, Clifton D, Ludbrook G, Young JD, Watkinson P: **The effect of fractional inspired oxygen concentration on early warning score performance: A database analysis**. *Resuscitation*. 2019; 139:192–199

*Joint first authors

Malycha J, Redfern O, Ludbrook G, Young JD, Watkinson PW. **Testing a digital system that ranks the risk of unplanned intensive care unit admission in all ward patients: Protocol for a prospective observational cohort study**. *BMJ Open* 2019; 9:1–4

Malycha J, Redfern OC, Pimentel M, Ludbrook G, Young JD, Watkinson PW. **Evaluation of a digital system to predict unplanned admissions to the intensive care unit: a mixed-methods approach**. *Submitted for publication*

Pimentel M, Redfern O, Malycha J, Meredith P, Prytherch D, Briggs J, Young JD, Clifton D, Tarassenko L, Watkinson PJ. **Detecting Deteriorating Patients in Hospital: Development**

and Validation of a Novel Scoring System. *American Journal of Respiratory and Critical Care Medicine*. 2021 Jul 1;204(1):44-52

Malycha J, Andersen C, Redfern O, Peake S, Phillips A, Subbe C, Ludbrook G, Young JD, Watkinson PJ, Flabouris A, Jones D. **Establishing consensus definitions for the deteriorating ward patient – a protocol.** *Submitted for publication*

Malycha J, Murphy D, Barker G, Ludbook G, JD Young, P Watkinson. **Using real-time location devices (RTLD) to quantify off-unit adult intensive care registrar workload: a 1-year tertiary NHS hospital prospective observational study**. *Journal of Clinical Monitoring and Computing*. 2019; 3–7

Thesis Signatures

Logistic challenges when collecting signatures from co-authors situated around the world (during the COVID-19 pandemic) meant Professor Guy Ludbrook (Principle Supervisor) signed on behalf of the contributing co-authors for all the publications included in the submitted thesis manuscript. All contributing co-authors did provide electronic signatures using the University of Adelaide Electronic Declaration Portal. These documents are stored by the University and the candidate and are available upon request.

Introduction

The research in this thesis describes the development of a system for the prediction of clinical deterioration in patients in the general hospital ward. This system is the Oxford University built Hospital-wide Alerts Via Electronic Noticeboard (HAVEN). The research was undertaken over five years, in a sequential manner and is presented in this thesis as published papers.

Rapid Response Systems (RRS) were first developed in the 1990's after critical care physicians in Australia, the United States (US) and the United Kingdom (UK) examined physiological trajectories in patients who suffered in-hospital cardiac arrest (IHCA) on general hospital wards. Their analysis showed IHCA was commonly heralded by prolonged periods (e.g., minutes to hours) of unnoticed physiological instability. (1–3) From these findings came the first Early Warning Scores (EWS), protocolised systems that scored abnormal vital signs and triggered responses when a threshold score was exceeded in patients on the general ward. (4, 5) From there, the RRS Afferent/Efferent Limb model was developed. EWS formed the 'Afferent Limb', the detection of deterioration with a trigger to initiate a response. The 'Efferent Limb' described the responding team, sometimes called the Medical Emergency Team (MET), a group of specialist clinicians trained in the management of critical illness. This linked RRS system was designed to optimise the allocation of specialist clinical resources to those most in need, thus reducing IHCA and preventable death in patients admitted to the general hospital ward. (6)

From 2000, RRS principles were incorporated into national clinical governance directives. The first International Conference on Medical Emergency Teams was held in the Pittsburgh, US, in 2005. (7) RRS implementation was stipulated in the Joint Commissions National Patient Safety Goals (the accreditation organisation for US hospitals) (2010), the Acute Medicine Task Force Report in the UK (2007) and the National Safety and Quality Health Service Standards in Australia (2012). (8)(9) In 2014 the formation of the International Society for Rapid Response Systems solidified the increasingly unified international approach to RRS research and development. Underpinning this evolution was a growing body of medical literature evaluating RRS. From 2000 onwards, EWS systems and validation studies proliferated. In 2013, the Royal College of Physicians of London facilitated the standardisation of EWS in the UK by developing the National Early Warning Score (NEWS). (10–14) Systematic reviews evaluating RRS efficacy supported the hypothesis that implementation of RRS were associated with a reduction in IHCA and death. (15–20)

Obtaining ‘Level 1’ evidence (i.e., evidence derived from blinded, randomised controlled trials) was more difficult. There are considerable methodological challenges associated with evaluating a ‘system based’ intervention like RRS. For example, randomising one patient to usual care whilst a neighbouring patient received care from the highly specialised MET raised both ethical and methodological issues. Despite these challenges, two randomised controlled trials were conducted using blocks of patients and step wise RRS introduction. In 2004, Priestley et al. published findings on the introduction of RRS via a ward (cluster) randomised method to a single UK hospital and found RRS reduced mortality. (21) The 2005 Modified Early Warning Score and the criteria used in the Medical Early Response Intervention and Therapy (MERIT) trial was the only multicentre, randomised control trial evaluating the implementation of MET criteria. This cluster study assigned 23 Australian hospitals to either continue functioning ‘as usual’ (n = 11), detecting, and treating deteriorating ward patients without RRS, or implement a Medical Emergency Team (n = 12). A composite primary outcome of IHCA, death and unplanned ICU admission was examined. MERIT demonstrated no significant difference in primary outcome between the control and intervention arms (5.86 v 5.31 events per 1000 admissions). (22) There were potential confounding factors effecting both trials, specifically the Hawthorne effect. Contamination between wards and hospitals affected ‘standard care’ in both studies, consequently reducing overall mortality.

Looking at the UK hospital population, despite the gains made using RRS over the last 30 years, untreated clinical deterioration in hospitalised ward patients persists. (23) Roughly 45,000 patients a year in hospital wards require urgent admission to an ICU. (24) Late recognition of this deterioration is associated with worse clinical outcomes, including higher mortality (25–27). Analysis of the UK Intensive Care National Audit and Research Centre (ICNARC) database shows up to 80% of unplanned ICU admissions from the wards have a preceding period of untreated clinical deterioration, despite the implementation of the National Early Warning Score in more than 75% of National Health Service Hospitals. (28)

The challenge for RRS has evolved from the detection and treatment of advanced to developing states of physiological deterioration. This evolution raises three hypotheses: Firstly, is it possible to detect and interpret subtle and early negative changes in the physiological status of ward patients, such that a clinical intervention (like RRS) mitigates the risk of subsequent and significant clinical deterioration? Secondly, can this process be augmented to the point that

accurate *detection* may be substituted by accurate *prediction*? Thirdly, can these hypotheses be evaluated in a meaningful way?

Between 2005 and 2105 increasing numbers of hospitals implemented Electronic Patient Records (EPR), thus creating opportunities to test these hypotheses by the implementation of advanced digital risk algorithms; (29)(30) algorithms that use rich and contemporaneous patient data sets including vital signs, laboratory results, demographics, comorbidities, and hospital administrative data. In 2008, Tam et al. developed a nomogram based on available administrative patient data to predict unplanned ICU admission but neither validated it nor published any follow up analysis. (31) Bailey et al. derived and implemented an EPR based prognostic risk score. (32) They found patients with a positive alert had a 5.3 times higher likelihood of needing ICU admission and had a significantly longer hospital length of stay. However, when implemented there was no difference in clinical outcomes of the intervention (EPR based alert ward patients) and control groups (standard RRS ward patients). In 2013, Alvarez et al. developed and validated an automated model that was more sensitive, specific and with a better Area Under the Receiver Operator Curve Characteristic (AUROC) 0.85 (95% CI 0.82 - 0.87), than contemporary systems. The model also identified patients destined to have adverse events on average 16 hours before they occurred. (33)

More recently, two international research groups have emerged as leaders in this field. Churpek et al., in Chicago USA, developed a predictive algorithm called the electronic Cardiac Arrest Risk Triage score (eCART). (34) eCART has been evaluated in multiple retrospective validation studies against widely used EWS, including Between The Flags (BTF), NEWS and the Modified Early Warning Score (MEWS) and was shown to better predict IHCA, unplanned ICU admission and death. (35) It has also been tested in a prospective feasibility study however this did not involve using the algorithm to alter current RRS clinical workflows. They demonstrated eCART detected four times as many IHCAs and 50% more ICU admissions than their current RRS. eCART also demonstrated detection of deterioration 8 hours in advance of the standard RRS activation thresholds. (34, 36) Escobar et al, in Oakland CA, are the most advanced team regarding real-world deployment of a predictive algorithm, which in their hospitals is called the Advanced Alert Monitor (AAM). They developed an intervention program where remote nurses reviewed the medical notes, and implemented a care plan, in patients identified by AAM as high risk. This system was implemented in a staggered fashion into 19 hospitals in the US and showed a reduction in mortality (Adjusted Relative Risk 0.84,

95% CI 0.78-0.90; p<0.001). (37) This study was the culmination of multiple sequential studies describing the development, validation and implementation of AAM. (38–43)

The Hospital-wide Alerts Via Electronic Noticeboard (HAVEN) is the University of Oxford equivalent. HAVEN was developed using supervised machine learning methods (gradient boosted machines) and used 76 routinely available, EPR-based variables. (44) These data were divided into two categories: static variables that do not change during an admission (e.g., age, past medical history), and dynamic variables that do change during a patient's admission and are repeatedly updated (e.g., vital signs, laboratory results). A risk estimate of deterioration in the next 24 hours was derived from these data. The risk estimate was then used to rank hospital ward patients from most to least likely to deteriorate in the coming 24 hours. This thesis describes the development of HAVEN.

References

1. Schein RMH, Hazday N, Pena M, et al.: Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest* 1990; 98:1388–1392
2. Smith AF, Wood J: Can some in-hospital cardio-respiratory arrests be prevented? A prospective survey. *Resuscitation* 1998; 37:133–137
3. Buist MD, Jarmolowski E, Burton PR, et al.: Recognising clinical instability in hospital patients before cardiac arrest or unplanned admission to intensive care. A pilot study in a tertiary-care hospital. *Med J Aust* 1999; 171:22–25
4. Stenhouse C, Coates S, Tivey M, et al.: Prospective evaluation of a modified Early Warning Score to aid earlier detection of patients developing critical illness on a general surgical ward. *Br J Anaesth* 2000; 84:663P
5. Subbe CP, Kruger M, Rutherford P, et al.: Validation of a modified Early Warning Score in medical admissions. *QJM* 2001; 94:521–526
6. Maharaj R: Textbook of Rapid Response Systems. 2018.
7. Devita M, Bellomo R, Hillman K, et al.: Findings of the first consensus conference on medical emergency teams. *Crit Care Med* 2006; 34:2463–2478
8. RCP: Acute medical care - The right person, in the right setting – first time. Report of the Acute Medicine Task Force. 2007.
9. The Australian Commission on Safety and Quality in Health Care: National Safety and Quality Health Service Standards. 2012.
10. Prytherch DR, Smith GB, Schmidt PE, et al.: ViEWS-Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* 2010; 81:932–937
11. Smith GB, Prytherch DR, Meredith P, 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; 84:465–470
12. Eccles SR, Subbe C, Hancock D, et al.: CREWS: Improving specificity whilst maintaining sensitivity of the National Early Warning Score in patients with chronic hypoxaemia. *Resuscitation* 2014; 85:109–111
13. Watkinson PJ, Pimentel MAF, Clifton DA, et al.: Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data. *Resuscitation* 2018; 129:55–60
14. Fernando SM, Fox-Robichaud AE, Rochwerg B, 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; 23:60

15. Winters BD, Pham JC, Hunt EA, et al.: Rapid response systems: A systematic review. *Crit Care Med* 2007; 35:1238–1243
16. Smith GB, Prytherch DR, Schmidt PE, et al.: Review and performance evaluation of aggregate weighted "track and trigger" systems. *Resuscitation* 2008; 77:170–179
17. Alam N, Hobbelink EL, van Tienhoven AJ, et al.: The impact of the use of the Early Warning Score (EWS) on patient outcomes: A systematic review. *Resuscitation* 2014; 85:587–594
18. Smith MEB, Chiovaro JC, O'Neil M, et al.: Early warning system scores for clinical deterioration in hospitalized patients: A systematic review. *Ann Am Thorac Soc* 2014; 11:1454–1465
19. Maharaj R, Raffaele I, Wendon J: Rapid response systems: A systematic review and meta-analysis. *Crit Care* 2015; 19:1–15
20. Fu LH, Schwartz J, Moy A, et al.: Development and validation of early warning score system: A systematic literature review. *J Biomed Inform* 2020; 105:103410
21. Priestley G, Watson W, Rashidian A, et al.: Introducing Critical Care Outreach: A ward-randomised trial of phased introduction in a general hospital. *Intensive Care Med* 2004; 30:1398–1404
22. Chen J, Flabouris A, Bellomo R, et al.: The Medical Emergency Team System and not-for-resuscitation orders: results from the MERIT study. *Resuscitation* 2008; 79:391–397
23. Hogan H, Hutchings A, Wulff J, et al.: Type of Track and Trigger system and incidence of in-hospital cardiac arrest: an observational registry-based study. *BMC Health Serv Res* 2020; 20:885
24. Intensive Care National Audit and Research Centre (ICNARC): Key statistics from the Case Mix Programme 2018/2019. *Case Mix Program Summ Stat* 2019; 0:1–6
25. Barwise A, Thongprayoon C, Gajic O, et al.: Delayed Rapid Response Team Activation Is Associated With Increased Hospital Mortality, Morbidity, and Length of Stay in a Tertiary Care Institution. *Crit Care Med* 2016; 44:54–63
26. National Confidential Enquiry into Patient Outcomes and Death: Time to Intervene? A review of patients who underwent cardiopulmonary resuscitation as a result of an in-hospital cardiorespiratory arrest. 2012.

27. Keogh B: Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report. *NHS* 2013;
28. Royal College of Physicians of London.: National Early Warning Score (NEWS) : standardising the assessment of acute-illness severity in the NHS. 2012.
29. Sendak MP, Arcy JD, Kashyap S, et al.: A Path for Translation of Machine Learning Products into Healthcare Delivery. 2020;
30. Adler-Milstein J, Holmgren AJ, Kralovec P, et al.: Electronic health record adoption in US hospitals: The emergence of a digital "advanced use" divide. *J Am Med Informatics Assoc* 2017; 24:1142–1148
31. Tam V, Frost SA, Hillman KM, et al.: Using administrative data to develop a nomogram for individualising risk of unplanned admission to intensive care. *Resuscitation* 2008; 79:241–248
32. Bailey TC, Chen Y, Mao Y, et al.: A trial of a real-time Alert for clinical deterioration in Patients hospitalized on general medical wards. *J Hosp Med* 2013; 8:236–242
33. Alvarez CA, Clark CA, Zhang S, et al.: Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. *BMC Med Inform Decis Mak* 2013; 13:28
34. Churpek MM, Yuen TC, Edelson DP: Risk stratification of hospitalized patients on the wards. *Chest* 2013; 143:1758–1765
35. Green M, Lander H, Snyder A, 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; 123:86–91
36. Kang MA, Churpek MM, Zadravecz FJ, et al.: Real-time risk prediction on the wards: A feasibility study. *Crit Care Med* 2016; 44:1468–1473
37. Escobar GJ, Liu VX, Schuler A, et al.: Automated Identification of Adults at Risk for In-Hospital Clinical Deterioration. *N Engl J Med* 2020; 383:1951–1960
38. Escobar GJ, Laguardia JC, Turk BJ, et al.: Early detection of impending physiologic deterioration among patients who are not in intensive care: Development of predictive models using data from an automated electronic medical record. *J Hosp Med* 2012; 7:388–395
39. Liu V, Turk BJ, Ragins AI, et al.: An electronic simplified acute physiology score-based risk adjustment score for critical illness in an integrated healthcare system. *Crit Care Med* 2013; 41:41–48

40. Escobar GJ, Turk BJ, Ragins A, et al.: Piloting electronic medical record–based early detection of inpatient deterioration in community hospitals. *J Hosp Med* 2016; 11:S18–S24
41. Escobar GJ, Dellinger RP: Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med* 2016; 11:S5–S10
42. Granich R, Sutton Z, Kim YS, et al.: Early detection of critical illness outside the intensive care unit: Clarifying treatment plans and honoring goals of care using a supportive care team. *J Hosp Med* 2016; 11:S40–S47
43. Paulson SS, Dummett BA, Green J, et al.: What Do We Do After the Pilot Is Done? Implementation of a Hospital Early Warning System at Scale. *Jt Comm J Qual Patient Saf* 2020; 46:207–216
44. Pimentel MA, Redfern OC, Malycha J, et al.: Detecting Deteriorating Patients in Hospital: Development and Validation of a Novel Scoring System. *Am J Respir Crit Care Med* 2021;

Chapter 1. Evaluating variables for potential inclusion in HAVEN

Prognostic algorithms that use EPR data to predict death, cardiac arrest and/or unplanned ICU admission in hospitalised patients have been developed over recent years. The models use a variety of statistical methods, but each requires patient derived 'candidate' variables (such as vital signs or laboratory results) to form the component parts of the model. The process of selecting candidate variables requires parsimony to avoid overfitting and numerous methods have been used, including expert opinion, forward and backward stepwise selection and more modern machine learning techniques. Evaluating univariate associations of candidate variables with the outcome of interest is the method described here. Univariate filter methods rank the strength of the statistical association. This is a pragmatic method when developing risk scores for hospital wide populations because it enables meaningful evaluation of the large numbers of candidate variables available within the EPR prior to their potential inclusion. This chapter describes the protocol and systematic review that provide a complete summary of patient derived variables with a univariate association with unplanned ICU admission. The measured time-dependent variables that were strongly associated with ICU admission were increased heart rate, respiratory rate, temperature and decreased systolic and diastolic blood pressure and arterial oxygen saturations. Additionally, increasing age, being male, a history of heart failure or diabetes and a diagnosis of hepatic disease have strong weight of evidence for an association. The lack of high-quality data in this field suggests further work is needed to establish the evidence base around when ICU admission is required.

VARIABLES ASSOCIATED WITH UNPLANNED GENERAL ADULT ICU ADMISSION IN HOSPITALISED PATIENTS: PROTOCOL FOR A SYSTEMATIC REVIEW

Authors

James Malycha MBBD FCICM[1,2,3], Tim Bonnici MD PhD[4], Katerina Sebakova MBBS FRACA [1], Tatjana Petrinic MSc[5], Duncan Young BM DM FRCA FMedSci[1], Peter Watkinson MD ChB MRCP FDICM[1]

Affiliations

[1]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[2]Discipline of Acute Care Medicine, University of Adelaide, South Australia, Australia
[3]Intensive Care Unit, The Queen Elizabeth Hospital, Adelaide, South Australia, Australia
[4]University College London Hospital, Intensive Care Unit, London, UK
[5]Bodleian Health Care Libraries, University of Oxford, Oxford, UK

Keywords

Clinical deterioration, critical care unit, electronic patient record, intensive care unit

Funding

This publication presents independent research funded by the Health Innovation Challenge Fund (HICF-R9-524; WT-103703/Z/14/Z), a parallel funding partnership between the Department of Health and Wellcome Trust.

Statement of Authorship

Title of paper	Variables associated with unplanned general adult ICU admission in hospitalised patients: Protocol for a systematic review
Publication status	Published
Publication details	Malycha J, Bonnici T, Sebekova K, Petrinic T, Young JD, Watkinson PJ: Variables associated with unplanned general adult ICU admission in hospitalised patients: Protocol for a systematic review. Syst Rev 2017; 6:4–7

Principal Author

Name of Principal Author	Dr James Malycha
Contribution to paper	Conceptualised the work, wrote manuscript and was corresponding author
Overall percentage (%)	90%
Certification	This paper reports on original research conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Co-author contributions

Each author certifies:

- The candidate's stated contribution to the publication is accurate (as state above);
- Permission is granted for the candidate to include the publication in the thesis; and
- The sum of all co-author contributions is equal less the candidate's stated contribution

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Tim Bonnici
Contribution to paper	Assisted with conceptualisation. Evaluated and edited the manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Katerina Sebakova
Contribution to paper	Evaluated and edited the manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor David Clifton
Contribution to paper	Evaluated and edited the manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Ms Tatjana Petrinic
Contribution to paper	Assisted literature search and methodology

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Duncan Young
Contribution to paper	Evaluated and edited the manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Peter Watkinson
Contribution to paper	Evaluated and edited manuscript

Abstract

Background

Failure to promptly identify deterioration in hospitalised patients is associated with delayed admission to intensive care units (ICUs) and poor outcomes. Existing vital-sign based Early Warning Score (EWS) algorithms do not have a sufficiently high positive predictive value to be used for automated activation of an ICU outreach team. Incorporating additional patient data might improve the predictive power of EWS algorithms however it is currently not known which patient data (or variables) are most predictive of ICU admission. We describe the protocol for a systematic review of variables associated with ICU admission.

Methods/Design

MEDLINE, EMBASE, CINAHL and the Cochrane Library, including Cochrane Database of Systematic Reviews and the Cochrane Central Register of Controlled Trials (CENTRAL) will be searched for studies that assess the association of routinely recorded variables associated with subsequent unplanned ICU admission. Only studies involving adult patients admitted to general ICUs will be included. We will extract data relating to the statistical association between ICU admission and predictor variables, the quality of the studies and the generalisability of the findings.

Discussion

The results of this review will aid the development of future models which predict the risk of unplanned ICU admission.

Systematic review registration: PROSPERO: CRD42015029617

Background

Unplanned admission of a patient in an acute care hospital to an intensive care unit (ICU) is a frequent occurrence. [1] In the UK, 40,000 patients per year have an unplanned Intensive Care Unit (ICU) admission from the general ward with up to 80% experiencing a preceding period of unchecked clinical deterioration. [2][3][4][5] Their mortality risk is 1.5 times higher than that of patients admitted directly to ICU from the emergency department. [2] Timely admission to an ICU may improve outcomes for these patients. [6] Many institutions worldwide use risk scores to trigger escalation in care. Escalation of care based on an Early Warning Score (EWS) is mandated in the UK. [7]

Despite implementation of EWS systems, missed clinical deterioration remains a significant problem. [8] Cognitive errors and barriers to communication have been identified as causes of missed deterioration. [9] In an attempt to bypass these problems some institutions have trialled directly linking electronic vital signs charts to alerting systems. [10] However, existing EWS algorithms, which are typically based on vital signs, have a poor positive predictive value for severe deterioration. [11] Therefore they cannot be usefully deployed in systems which aim to automatically alert trained specialists to impending deterioration on the ward as the number of false alerts is excessive. Inclusion of additional variables can improve the accuracy of EWS models. [12][13]

Objective

We will conduct a systematic review to identify studies of patient-derived variables that are associated with an increased risk of unplanned ICU admission. For the purposes of the review, a variable is defined as an indivisible entity, as opposed to a composite entity such as a risk score, which is made up of multiple variables. A patient-derived variable is a measure of the properties of a patient as opposed to a measure of institutional processes such as nurse-to-patient ratio or number of escalation calls.

Methods /Design

This protocol will adhere to the requirements of Preferred Reporting Items for Systematic Reviews and Meta-analyses Protocol (PRISMA-P), which is included as Additional File 1.

Search Strategy

Papers will be identified by searching Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica database (EMBASE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), the Cochrane Database of Systematic Reviews and the Cochrane Central Register of Controlled Trials (CENTRAL). We will include additional papers from the references of reviews articles or studies identified during screening and papers from the authors' personal libraries. A full description of the search strategy is outlined in Appendix 1.

Study Selection

Two researchers will independently screen titles and abstracts of identified papers against the inclusion and exclusion criteria. They will not be blinded to the journal titles or to the study authors or institutions. If there is disagreement or uncertainty regarding eligibility, the article will be included in the next stage of screening for further analysis for inclusion/exclusion. The full text will be retrieved for all articles not excluded by the initial screening. These papers will be independently assessed against the inclusion and exclusion criteria. Disagreements about eligibility will be resolved by discussion between the screening researchers or a third party.

Inclusion criteria

Types of studies

Quantitative studies published in peer reviewed journals assessing adults admitted to adult hospitals will be eligible for inclusion in this review. Studies will most likely be prospective or retrospective cohort and case-control studies.

Study Characteristics

Eligible studies must include both a cohort of patients admitted to ICU and a cohort not admitted to ICU. Unplanned ICU admission may be either a primary or secondary outcome measure. Studies published from January 2000 until the day of search completion will be included to ensure modern day applicability. No language restrictions will be applied.

Phenomenon of Interest

Studies must describe a statistical relationship between a patient-derived variable (e.g., heart rate or creatinine level) and an unplanned admission to intensive care from a general ward or

emergency department. ‘Diagnosis’ or ‘groups of diagnoses’ are eligible to be included as variables. If a paper analyses both eligible variables (e.g., variables that are widely available in most UK hospitals) and non-eligible variables (e.g., variables that are not widely available in most UK hospitals) it will still be eligible for inclusion, with the authors using only the eligible variables for inclusion in the review.

Population

Studies that sample adult patients with an unplanned admission to ICU will be considered for inclusion. For the purpose of this review adult is defined as > 16 years of age. There will be no other restrictions.

Exclusion criteria

Types of studies

Qualitative studies, case studies, grey-literature, editorials, letters, practice guidelines and abstract-only reports will be excluded.

Study Characteristics

Studies of cohorts defined by a single condition or narrow group of conditions (e.g., trauma or sepsis) will be excluded. We will also exclude studies that do not use a control versus intervention group.

Phenomenon of Interest

Studies of ICU readmission or admission to ICUs dedicated to narrow cohorts of patients will be excluded (e.g., patients admitted to ICU with acute liver failure).

Population

Studies of participants under 16 years old will be excluded.

Data Extraction

Two authors will independently extract data from the papers and supplementary material. All uncertainties regarding data extraction will be resolved by discussion amongst the study team. DistillerSR (Evidence Partners, Ottawa, Canada) will be used to manage the data and identify

duplicate search results. All screening and data extraction forms will be implemented within DistillerSR. As part of the development of this protocol the study forms have been piloted and a calibration exercise has been undertaken to ensure good inter-rater agreement.

Quality Assessment

Risk of bias will be assessed using a scoring system adapted from two previous systematic reviews, [14][15] both of which are adapted from the Newcastle-Ottawa Scale (NOS). [16] The NOS is a scoring system designed to assess the quality of nonrandomised studies in meta-analyses. Using a 'star' system, it attributes a score to a paper after assessing the selection of study groups, the comparability of the groups and the ascertainment of either the exposure or outcome of interest for case-control or cohort studies. The scoring system used in this systematic review is outlined in Appendix 2.

Data Synthesis and analysis

We will conduct a qualitative synthesis of results from included studies. This will be presented descriptively in table and text form. We will extract summary comparison data as odds ratios (95% confidence intervals) where possible. Where sufficient original data is presented, we will calculate odds ratios. Where insufficient data is presented, we will contact the authors.

Discussion

As hospitals move towards fully digital patient records, increasing amounts of data are being collected in hospital Clinical Information Systems. Researchers have begun using this resource to develop models to predict patient deterioration based upon electronically captured data. [17,18] These models are reported to perform better than conventional EWS algorithms but their clinical adoption is not widespread.

Commonly, patient deterioration prediction models aim to accurately predict one of cardiac arrest, death or unplanned ICU admission. This systematic review will be the first to bring together the hospitalised patient factors that are known to be associated with subsequent urgent admission to ICU alone. This is a vital step in starting to use this information to identify patients at risk of ICU admission.

The findings from this review will contribute to the construction of an improved model for the prediction of clinical deterioration and unplanned ICU admission in adult patients on general wards. The findings may also be useful for researchers seeking to improve upon existing work in this field.

List of Abbreviations

CENTRAL - The Cochrane Database of Systematic Reviews and the Cochrane Central Register of Controlled Trials
CINAHL - Cumulative Index to Nursing and Allied Health Literature
EMBASE - Excerpta Medica database
EWS – Early Warning Score
ICU – Intensive Care Unit
MEDLINE - Medical Literature Analysis and Retrieval System Online
NICE – National Institute of Clinical and Healthcare Excellence

References

1. The Health and Social Care Information Centre. Adult Critical Care in England - April 2010 to March 2011: Experimental Statistics. Hospital Episode Statistics; 2012 May pp. 1–43.
2. UK Intensive Care National Audit Registry National Report 2012/13
3. Kause J, Smith GB, Prytherch DR, Parr M, Flabouris A, Hillman KM, 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;62:275–82.
4. Stelfox HT, Bagshaw SM, Gao S. Characteristics and Outcomes for Hospitalized Patients With Recurrent Clinical Deterioration and Repeat Medical Emergency Team Activation. Crit. Care Med. 2014;42:1601–9.
5. Buist MD, Bernard S, Nguyen TV, Moore G, Anderson J. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. Resuscitation. 2004;62:137–41.
6. Liu V, Kipnis P, Rizk NW, Escobar GJ. Adverse outcomes associated with delayed intensive care unit transfers in an integrated healthcare system. J. Hosp. Med. Wiley Subscription Services, Inc., A Wiley Company; 2012;7:224–30.
7. National Institute of Clinical Excellence. Acutely ill patients in hospital. NICE, editor. 2007 Jul pp. 1–107. Report No.: NCG50.
8. Findlay GP, Shotton H, Kelly K, Mason M. Time to Intervene? NCEPOD. National Confidential Enquiry into Patient Outcome and Death (NCEPOD); 2012.
9. Shearer B, Marshall S, Buist MD, Finnigan M, Kitto S, Hore T, et al. What stops hospital clinical staff from following protocols? An analysis of the incidence and factors behind the failure of bedside clinical staff to activate the rapid response system in a multi-campus Australian metropolitan healthcare service. BMJ Qual Saf. 2012;21:569–75.
10. Jones S, Mullally M, Ingleby S, Buist MD, Bailey M, Eddleston JM. Bedside electronic capture of clinical observations and automated clinical alerts to improve compliance with an Early Warning Score protocol. Crit Care Resusc. 2011;13:83–8.
11. Romero-Brufau S, Huddleston JM, Naessens JM, Johnson MG, Hickman J, Morlan BW, et al. Widely used track and trigger scores: are they ready for automation in practice? Resuscitation. 2014;85:549–52.

12. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Meltzer DO, Gibbons RD, et al. Multicenter Development and Validation of a Risk Stratification Tool for Ward Patients. American Journal of Respiratory and Critical Care Medicine. 2014;190:649–55.
13. Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous measure of patient condition using the Electronic Medical Record. Journal of Biomedical Informatics. 2013;46:837–48.
14. Taggart DP, D'Amico R, Altman DG. Effect of arterial revascularisation on survival: a systematic review of studies comparing bilateral and single internal mammary arteries. Lancet. 2001 Sep 15;358(9285):870-5.
15. Duckitt, D. Harrington, Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies. BMJ 330, 565 (2005)
16. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P: The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2013.
17. Churpek MM, Yuen TC, Park SY, Meltzer DO, Hall JB, Edelson DP. Derivation of a cardiac arrest prediction model using ward vital signs. Crit. Care Med. 2012;40:2102–8.
18. Escobar GJ, LaGuardia JC, Turk BJ, Ragins A, Kipnis P, Draper D. Early detection of impending physiologic deterioration among patients who are not in intensive care: Development of predictive models using data from an automated electronic medical record. J. Hosp. Med. 2012;7:388–95.

Appendix 1: Draft Search Strategy for MEDLINE

1.	(ICU* OR "intensive care" OR "critical care").ab,ti.
2.	INTENSIVE CARE UNITS/
3.	1 OR 2
4.	(admission* OR admitted OR transfer*).ab,ti.
5.	("risk assessment*" OR "risk factor*" OR "risk stratif*" OR predict* OR "increased risk*" OR trigger* OR score* OR scoring OR "early warning" OR escalat* OR deteriorat* OR triag* OR "vital sign*" OR model* OR validat*).ab,ti.
6.	3 AND 4 AND 5
7.	limit 6 to (humans AND yr="2000 -Current")
8.	(observational OR "case control*" OR retrospective OR cohort* OR "systematic review*").ab,ti.
9.	OBSERVATIONAL STUDY/
10.	CASE-CONTROL STUDIES/
11.	RETROSPECTIVE STUDIES/
12.	COHORT STUDIES/
13.	RANDOMIZED CONTROLLED TRIAL/
14.	REVIEW/
15.	COMPARATIVE STUDY/
16.	PROSPECTIVE STUDIES/
17.	VALIDATION STUDIES/
18.	8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17
19.	7 AND 18
20.	(unplanned OR unexpected OR unanticipated OR emergency OR "rapid response").ab,ti.
21.	19 AND 20

Appendix 2: Risk of bias scoring

Participant Selection	**Score**
Cohort studies	
Selected cohort is very representative of the general hospitalised population	**2**
Selected cohort is somewhat representative of the general hospitalised population	**1**
Cohort is not representative of the general hospital population or the selection of the group was not described	**0**
Case-control studies	
Cases and controls drawn from the same population and population is very representative of the general hospitalised population	**2**
Cases and controls drawn from the same population and population is somewhat representative of the general hospitalised population	**1**
Cases and controls drawn from different sources or the selection of groups was not described	**0**
Comparability of groups	
No differences between the groups explicitly reported unless it was one of these	**2**
variables that was under investigation, or such differences were adjusted for	**1**
Differences between groups were not recorded	**0**
Groups differed	
Size	
> 100 participants in each group	**2**
< 100 participants in each group	**1**
Adjustment for confounding	**1**
Adjustment made for confounding factors in data analysis	**1**
No adjustment for cofounders	**0**

PATIENT CENTRED VARIABLES WITH UNIVARIATE ASSOCIATIONS WITH UNPLANNED ICU ADMISSION: A SYSTEMATIC REVIEW

Authors

James Malycha MBBD FCICM[1,2,3], Tim Bonnici MD PhD[4], David A. Clifton MEng DPhil (Oxon) CEng[5], Guy Ludbrook PhD MBBS FANZCA[2], Duncan Young BM DM FRCA FMedSci[1], Peter Watkinson MD ChB MRCP FDICM[1]

Affiliations

[1]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[2]Discipline of Acute Care Medicine, University of Adelaide, South Australia, Australia
[3]Intensive Care Unit, The Queen Elizabeth Hospital, Adelaide, South Australia, Australia
[4]University College London Hospital, Intensive Care Unit, London, UK
[5]Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK

Keywords

Critical care; intensive care; ICU admission; clinical deterioration; EPR; EHR; variable selection; systematic review; predictive scores

Funding

This publication presents independent research funded by the Health Innovation Challenge Fund (HICF-R9-524; WT-103703/Z/14/Z), a parallel funding partnership between the Department of Health and Wellcome Trust.

Statement of Authorship

Title of paper	Patient centred variables with univariate associations with unplanned ICU admission: A systematic review
Publication status	Published
Publication details	Malycha J, Bonnici T, Clifton DA, Ludbrook G, Young JD, Watkinson PJ: Patient centred variables with univariate associations with unplanned ICU admission: A systematic review. BMC Med Inform Decis Mak 2019; 19:4–7

Principal Author

Name of Principal Author	Dr James Malycha
Contribution to paper	Conceptualised the work, wrote manuscript and was corresponding author
Overall percentage (%)	90%
Certification	This paper reports on original research conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Co-author contributions

Each author certifies:

- The candidate's stated contribution to the publication is accurate (as state above);
- Permission is granted for the candidate to include the publication in the thesis; and
- The sum of all co-author contributions is equal less the candidate's stated contribution

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Tim Bonnici
Contribution to paper	Conceptualised work, evaluated and edited the manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor David Clifton
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Guy Ludbrook
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Duncan Young
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Peter Watkinson
Contribution to paper	Conceptualised work, evaluated and edited manuscript

Abstract

Background

Multiple predictive scores using Electronic Patient Record data have been developed for hospitalised patients at risk of clinical deterioration. Methods used to select patient centred variables for inclusion in these scores varies. We performed a systematic review to describe univariate associations with unplanned Intensive Care Unit (ICU) admission with the aim of assisting model development for future scores that predict clinical deterioration.

Methods

Data sources were MEDLINE, EMBASE, CINAHL, CENTRAL and the Cochrane Database of Systematic Reviews. Included studies were published since 2000 describing an association between patient centred variables and unplanned ICU admission determined using univariate analysis. Two authors independently screened titles, abstracts and full texts against inclusion and exclusion criteria. DistillerSR (Evidence Partners, Canada, Ottawa, Ontario) software was used to manage the data and identify duplicate search results. All screening and data extraction forms were implemented within DistillerSR. Study quality was assessed using an adapted version of the Newcastle-Ottawa Scale. Variables were analysed for strength of association with unplanned ICU admission.

Results

The database search yielded 1520 unique studies; 1462 were removed after title and abstract review; 57 underwent full text screening; sixteen studies were included. One hundred and eighty nine variables with an evaluated univariate association with unplanned ICU admission were described.

Discussion

Being male, increasing age, a history of congestive cardiac failure or diabetes, a diagnosis of hepatic disease or having abnormal vital signs were all strongly associated with ICU admission.

Conclusion

These findings will assist variable selection during the development of future models predicting unplanned ICU admission.

Background

In experimental settings, scores that predict risk for clinical deterioration in hospitalised patients have evolved from vital sign based Early Warning Scores (EWS) to systems that utilise the large amount of patient centred data in Electronic Patient Records (EPRs).(1-4) These systems are not yet in widespread use, however they represent a first step towards *automatically* assimilating patient data to assist clinical decision making on high risk ward patients. Each of the current, published experimental models were derived and validated on large EPR linked databases that used Intensive Care Unit (ICU) admission as one of the outcome measures. This outcome measure is commonly used (along with death and cardiac arrest) as a surrogate for confirmed clinical deterioration.

These and other prognostic models use a variety of statistical methods but multivariate regression modelling and machine learning techniques are commonly used. These methods require patient centred 'candidate' variables (such as vital signs or laboratory results) to form the component parts of the model.(5) The process of selecting model candidate variables is important, however there is no consensus on how best to do this. Numerous methods have been used for multivariate logistic regression, including expert opinion, forward and backward stepwise selection and machine learning techniques.(6) A logical and often used first step is evaluating univariate associations, which enables the variables to be quantified in advance of their inclusion in the model.(7) This is helpful when using EPR data where there are large number of available candidate variables.(8) Regardless of the method, the goal is to include the *optimal* combination of variables that maximise predictive ability, whilst avoiding unnecessary complexity.(6)

In this systematic review we provide a complete summary of patient centred variables with a *univariate* association with unplanned ICU admission. By providing these data, we hope to aid the development of EPR based models for the prediction of ICU admission (and therefore clinical deterioration). We anticipate these data will enhance data-driven improvements in the care of deteriorating ward patients.

Methods

Search and Identification of Studies

The study protocol has been published (9) and follows the Preferred Reporting of Observational Studies and Meta-Analysis (PRISMA) statement. (10) An experienced medical librarian helped devise the search strategy to maximise identification of relevant studies. Studies were identified by searching Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica database (EMBASE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), the Cochrane Database of Systematic Reviews and the Cochrane Central Register of controlled trials (CENTRAL). We included additional studies from the references of review articles, studies identified during screening, and from the authors' personal libraries. We restricted the search to studies published since 2000. We did not apply any language restrictions. The search design is shown in the **supplemental digital content (SDC-1)**.

Inclusion Criteria

Included studies evaluated hospitalised, adult patients located in either the Emergency Department (ED), the general surgical or medical wards. Patients in specialist wards (such as obstetric or psychiatric) were eligible if they were evaluated as a part of the general patient population rather than disease specific sub groups of patients. Included studies required an analysis of at least two cohorts: one cohort of patients admitted to ICU (intervention) and one not admitted to ICU (control). Variables were eligible if they were patient centred and had been evaluated across both cohorts. Studies which described a univariate, statistical relationship between a patient centred variable (e.g., heart rate) and unplanned ICU admission were included. The described variables were single entities, as opposed to composites such as risk scores.

Exclusion Criteria

Excluded studies did not evaluate unplanned ICU as an isolated outcome measure nor did they evaluate patients requiring ICU readmission. Additionally, studies that evaluated variables related to hospital processes or environmental risk alone (e.g., staff-to-nurse ratios), carried out multivariate analyses (without describing the univariate analyses that went into selecting variables for the model) or evaluated patient groups with a single diagnosis, were also excluded. (The studies excluded via this criterion are listed in **supplementary digital content**

(SDC-2)). Patients admitted to ICU (or not) from high acuity areas such as HDU were excluded from the review as these hospital areas are often linked to ICUs and so are not always captured as admissions. Subgroups of illness acuity, such as needing an Rapid Response System (RRS) alert or being a high triage category, were not excluded. Finally, studies not published in peer-reviewed journals and those examining patients < 15 years old were excluded.

Study Selection and Data Abstraction

Two authors (JM, TB) independently screened titles and abstracts of identified studies against the inclusion and exclusion criteria. **(Figure 1)** They were not blinded to the journal titles or to the study authors or institutions. If there was disagreement or uncertainty regarding eligibility, the article was included in the next stage of screening. The full text was retrieved for all articles not excluded by the initial screening and re-assessed for eligibility as before. Disagreements about eligibility were resolved by discussion between the screening authors or a third party (a senior member of our research team PW and DY). Two authors extracted independently data from the studies and supplementary material. Any uncertainties regarding data extraction were resolved by discussion amongst the study team. DistillerSR (Evidence Partners, Canada, Ottawa, Ontario) was used to manage the data and identify duplicate search results. All screening and data extraction forms were implemented within DistillerSR.

Risk of Bias Assessment

Two authors (JM, TB) independently assessed the risk of bias for included studies by using an adapted version of the Newcastle-Ottawa Scale (NOS) (11, 12) The NOS is a scoring system designed to assess the quality of non-randomised studies in meta-analyses and systematic reviews. (13) We generated a score by assessing each study cohort for representativeness of the desired study population, the comparability of the cohorts being assessed, the size of the cohorts and correction for confounding. We adapted the NOS score to show bias in the types of studies included in this review (i.e. those showing univariate associations) whilst remaining faithful to the established NOS methodology. The details of the modified NOS scale are included in the **supplementary digital content (SDC-3)**.

Data Synthesis

We categorised patient-derived variables as comorbidities, demographics, laboratory tests, vital signs, diagnoses, medications and symptoms/signs, in general accordance with the categories used in the included studies. To synthesise and present the large number of variables included in the results in a logical way, we adopted the recently published method of Dettmer et al. (as adopted from Zaal et al.), who combined the quality of the studies investigating the variables in question (based on the NOS risk of bias assessment) with the number of times the variable was investigated. (14, 15) This semi-quantitative approach enable the assignation of a 'weight of evidence' to each variable **(Table 1)**.

Results

All included studies are shown in **Table 2**. (16-31) The database search yielded 1520 unique studies; 1462 were removed after title and abstract review; 57 underwent full text screening; sixteen studies were included in this review **(Figure 1)**. Summary details are shown in **Table 2** with additional study data in the **supplementary digital content (SDC-4).** The mean study quality score was five and the mode was seven. We graded six studies high quality (18, 20,21,28-30), four moderate quality (16, 23-25) and six (low quality). (17, 19, 22, 26, 27, 31) The results of the bias assessment for each study are shown in **Table 2** and the **supplementary digital content (SDC-5).** The quality of the studies is also reflected in the weight of evidence for any particular variable.

Quantised data were presented as independent variables. For example, arterial oxygen saturation was evaluated six times for ranges of <80%, <95%, 80-89%, 90-94%, mean % and median (%) (in each study group) across four studies and thus included six times in the initial analysis, with each of these ranges being defined as a single variable. (16, 18, 20, 22) Likewise, 'cardiovascular disorder' was included 14 times across six studies, as either a comorbidity, diagnosis or symptom/sign. (21, 26-29) We recorded the statistical relationship between variables and unplanned ICU admission as p values, Odds Ratios (OR), Risk Ratios (RR) (28) or Incidence Rate Ratios (IRR) (21) (with 95% confidence intervals) and did not assign preference.

Five studies (31%) were case control studies and eleven (69%) were cohort studies. Eight (50%) were prospective and eight (50%) were retrospective. The number of participants in each

study ranged between 95 in a prospective cohort study and 126,826 in a retrospective case control study. The number of patients admitted to ICU ranged between fifteen and 5233, while in the control group they ranged between 80 and 125,244 **(Table 2)**. Five studies (31%) evaluated patients in emergency departments (ED) and eleven studies (69%) evaluated patients treated on hospital wards. Of the studies examining ward patients, five evaluated patients admitted via ED, two evaluated patients who had an RRS review and four evaluated patients admitted via any source **(SDC-6, supplementary digital content).** Escobar (20) studied patients in twenty centres, Schuetz (25) in three centres and Loekiko (24) in two centres. The remaining thirteen studies were single centre **(Table 1)**

Across the sixteen studies, 189 different patient-derived variables were assessed for univariate association with unplanned ICU admission. Of these, 53 were vital signs, 42 were comorbidities, 29 were diagnosis, 26 were demographics, 25 were laboratory results, 10 were symptoms/signs and 4 were medications. One hundred and twenty-eight variables had a statistically significant positive association, two had a negative association and 59 had no association with unplanned ICU admission. Information on effect size was described as ORs, RRs or IRRs where available and is shown in **supplementary digital content (SCD-7).**

The semi-quantitative analysis resulted in 110 variables after repeatedly measured variables were grouped together. These are shown in **Table 3** and **supplementary digital content (SDC-8).** Overall there were twelve variables with a strong weight of evidence (one was negative), three with a moderate weight of evidence and 33 with a weak weight of evidence for an association with unplanned ICU admission. The remaining 62 variables showed an inconclusive weight of evidence.

Variables associated with unplanned ICU admission

Variables with a strong, moderate and negative weight of evidence for association with unplanned ICU admission are summarised in **Table 3.**

Comorbidities, demographics and diagnosis

A history of congestive heart failure and diabetes were the only comorbidities in this group. These had a significant result in two high-quality studies (21, 29) Of the demographics, being male (20, 21, 27, 28) and an increasing age (16, 18, 20, 21, 27, 28) had a strong weight of

evidence for association. Four studies showed a significant difference in mean or median age between ICU and control groups (higher in the ICU group) (18, 20, 21, 27) and three studies showed a significant OR or IRR for increased age quantiles with the oldest quantile being 75+ years of age. (21, 28) Hepatic disease was the only diagnosis strongly associated with unplanned ICU admission. (28, 30)

Vital signs

All six vital signs had a strong association. Heart rate was studied twelve times across seven studies, ten times as quantiles (seven times for tachycardia and three times for bradycardia) and twice as a comparison of means. (16, 18, 20, 22, 27, 29, 30) Seven of the tachycardia and one of the bradycardia quantiles (< 60 beats per minute) showed a positive association**.** Two high-quality studies also found a significant difference in mean heart rate (higher in the ICU group). (18, 20) Elevated respiratory rate was evaluated eight times across six studies and had a strong weight of evidence. (16, 20, 22, 27, 30) Five of the six quantiles showed a significant result (16, 27, 30) and both high-quality studies examining mean respiratory rate showed a significant difference. (18, 20) The only non-significant result was a respiratory rate of > 20 breaths per minute in a low-quality study. (22) Systolic blood pressure (SBP) was evaluated seven times across five studies. (16, 18, 20, 22, 27) Two high-quality studies showed a significant reduction in mean blood pressure (18, 20) and one moderate-quality study showed a significant OR for a SBP of 80-89mmHg versus 90mmHg and above. (16) Diastolic blood pressure (DBP) and temperature were evaluated in the same two high-quality studies, both showing significant differences in mean (decreased for SBP and increased for temperature). (18, 20) Both studies had very small variations, < 0.2^{o}Cand < 2mmHg respectively. Arterial oxygen saturation was studied six times across four studies. (16, 18, 20, 22) Lower saturation quantiles (<80%, 80-89% and 90-94%) and lower mean/median saturations were shown to be significant.

Variables moderately and weakly associated with unplanned ICU admission are summarised in **Table 3** and **supplementary digital content (SDC-8)** respectively.

Discussion

Statement of findings

In this systematic review of 16 observational and cohort studies evaluating ED and ward patients, we found two comorbidities (congestive cardiac failure and diabetes), two

demographics (increasing age and being male), one diagnosis (hepatic disease) and six vital signs (respiratory rate, heart rate, temperature, systolic and diastolic blood pressure and arterial oxygen saturations) with a strong univariate association with unplanned ICU admission. These findings support the consensus that abnormal vital signs have significant value when predicting unplanned ICU admission. The strength of association for a history of congestive cardiac failure and diabetes and a new diagnosis of hepatic disease may reflect the high burden of care required in this patient cohort up until the terminal phase of disease. Being older and male as a risk factor for ICU admission may reflect the general hospital population as a whole. Overall this review provides a thorough summary of the candidate variables available in EPRs (and elsewhere in the clinical record) that will assist researchers to develop and evaluate predictive models for patients at risk of unplanned ICU admission.

Clinical and research implications

Progressing from vital sign based, EWS systems to EPR based, risk model systems has incrementally improved performance, both in terms of correctly identified deteriorating ward patients (sensitivity) and the number of 'false alarms' generated for clinical staff (specificity and positive predictive value). These performance gains have been achieved via multivariate regression models and more recently machine learning processes. (1, 35, 36) Regardless of the statistical approach, candidate variables should be selected in a methodologically robust way. In the published literature, univariate filter methods, that rank the strength of the statistical association, are among the most common. (7, 8) It is a popular approach because the univariate analysis provides a summary of the variables most likely to enhance model performance, does not involve significant computation, is relatively simple, not time consuming and produces an easily interpretable output. It does have weaknesses however, including the potential to miss variables that have no association with the outcome when evaluated *in isolation* but have an association when evaluated together with another variable (e.g., age).

Despite their performance advantage, as yet no EPR based hospital model has achieved widespread adoption. In contrast, 75% of UK National Health Service hospitals monitor ward patients using the National Early Warning Score (NEWS) (37, 38) The success of NEWS, which is a simple aggregate score that uses the univariate associations of abnormal vital signs with adverse patient outcomes, highlights the importance of interpretability and generalisability in this research and clinical domain. Advanced scoring systems that rely on

complex computational processes may be difficult to interpret (and trust) for clinical staff and therefore less likely to be adopted into general use. We hope the univariate associations described will provide a convenient and intuitive reference for clinicians and researchers alike to overcome such barriers to implementation.

Strengths and limitations

The association of the variables does not infer causality. The search strategy was thorough and in accordance with current methodological guidelines but studies may have been missed. Publication bias may have affected results. The methodology of the included studies was varied, making meta-analysis inappropriate. We excluded studies examining specific sub-populations of patients only (i.e. acute liver failure) meaning the variables summarised in this review are not applicable for risk models designed for specific disease sub-groups.

There is a lack of consensus on which outcome measures to use when assessing the performance of predictive models for clinical deterioration. (18) Each of cardiac arrest, in hospital death and unplanned ICU admissions represent different populations and will, therefore, have different variable associations. We selected unplanned ICU admission as an isolated outcome measure (and excluded in hospital death and cardiac arrest) in the knowledge this would reduce the number of eligible studies and therefore potential variables for inclusion in this review. We adopted this method because we aim to advance the study of models that predict clinical deterioration, specifically in those who will most benefit from an intervention such as an ICU admission. When predicting ICU admission, some authors published the univariate relationships from within their derivation databases before including them in the multivariate analysis.(20) However we are not aware of any who have based selection on associations evaluated in external databases.

We evaluated a heterogeneous study population by including ED, ward, post-Medical Emergency Team (MET) and non-post-MET patients. This was done because we wish to better understand associations with clinical deterioration, which may occur at any time-point during hospital admission. Namely, early deterioration, which may occur soon after discharge from ED to the ward, in which case patient centred ED data is important. Or late deterioration, when the patient has been on the ward for some days. Studies examining sub-populations of patients (i.e. where specifically designed predictive algorithms have the potential to be more accurate

than when used in a general patient population) were excluded on the basis that as a first step, we wish to isolate variables that will contribute to a hospital wide EPR based risk score.

We have deliberately avoided describing multivariate studies because we do not wish describe the models themselves. There are multiple examples of high performing, multivariate clinical predictive models in the literature, whose variables will have quantifiable associations with unplanned ICU admission. However, it is impossible to exclude collinearity in these instances, making obsolete our objective to individually quantify these variable associations as potential "building blocks" for future models. As a consequence, multivariate analysis was excluded unless the univariate associations were described.

Conclusion

Having abnormal vital signs, being elderly, male, having a history of heart failure or diabetes and a diagnosis of liver failure are all strongly associated with unplanned ICU admission. This systematic review is the first to comprehensively collate the evidence on patient centred variables with univariate associations with ICU admission. These results may assist the development of predictive models for hospitalised patients at risk of needing escalations in care. There is a lack of high-quality data in this field and further work is required to isolate the patient centred variables most likely to enhance model performance when predicting unplanned ICU admission.

Abbreviations

CI Chief Investigator
CTRG Clinical Trials & Research Governance, University of Oxford
GCP Good Clinical Practice
REC Research Ethics Committee
RRS Rapid Response System
ICU Intensive Care Unit
EPR Electronic Patient Record
OR Odds Ratio
RR Risk Ratio
IRR Incidence Rate Ratios
NEWS National Early Warning Score

ED Emergency Department

References

1. Kipnis P, Turk BJ, Wulf DA, 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; 64:10–19
2. Bailey TCC, Chen Y, Mao Y, et al.: A trial of a real-time Alert for clinical deterioration in Patients hospitalized on general medical wards. J Hosp Med 2013; 8:236–242
3. Kang MA, Churpek MM, Zadravecz FJ, et al.: Real-Time Risk Prediction on the Wards. Crit Care Med 2016; 44:1468–1473
4. Alvarez CA, Clark CA, Zhang S, et al.: Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. BMC Med Inform Decis Mak 2013; 13:28
5. Steyerberg E, Moons KGM, van der Windt D, et al.: Prognosis research strategy (PROGRESS) series 3: prognostic model research. PLoS Med 2013; 10:e1001381
6. Sanchez-Pinto LN, Venable LR, Fahrenbach J, et al.: Comparison of variable selection methods for clinical predictive modeling. Int J Med Inform 2018; 116:10–17
7. Walter S, Tiemeier H: Variable selection: Current practice in epidemiological studies. Eur J Epidemiol 2009; 24:733–736
8. Bagherzadeh-Khiabani F, Ramezankhani A, Azizi F, et al.: A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results. J Clin Epidemiol 2016; 71:76–85
9. Malycha J, Bonnici T, Sebekova K, et al.: Variables associated with unplanned general adult ICU admission in hospitalised patients: Protocol for a systematic review. Syst Rev 2017; 6:4–7
10. Moher D, Liberati A, Tetzlaff J, et al.: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009; 6
11. Taggart DP, D'Amico R, Altman DG: Effect of arterial revascularisation on survival: A systematic review of studies comparing bilateral and single internal mammary arteries. Lancet 2001; 358:870–875
12. Duckitt K, Harrington D: Risk factors for pre-eclampsia at antenatal booking: Systematic review of controlled studies. Br Med J 2005; 330:565–567
13. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M et al. TN-OS (NOS) for assessing the quality of nonrandomised studies in meta-analyses. [cited 2016 A 25].

A from: http://www. ohri. ca/programs/clinical_epidemiology/oxford. asp.: Our Research. Crick Cent 2015; 2018

14. Dettmer MR, Damuth E, Zarbiv S, et al.: Prognostic Factors for Long-Term Mortality in Critically Ill Patients Treated with Prolonged Mechanical Ventilation: A Systematic Review. Crit Care Med 2017; 45:69–74
15. Zaal IJ, Devlin JW, Peelen LM, et al.: A Systematic Review of Risk Factors for Delirium in the ICU*. Crit Care Med 2015; 43:40–47
16. Barfod C, Lauritzen MMP, Danker JK, et al.: Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency department - a prospective cohort study. Scand J Trauma Resusc Emerg Med 2012; 20:28
17. Calzavacca P, Licari E, Tee A, et al.: Point-of-care testing during medical emergency team activations: A pilot study. Resuscitation 2012; 83:1119–1123
18. Churpek MM, Yuen TC, Edelson DP: Predicting clinical deterioration in the hospital: The impact of outcome selection. Resuscitation 2013; 84:564–568
19. Eick C, Rizas KD, Meyer-Zurn CS, et al.: Autonomic nervous system activity as risk predictor in the medical emergency department: a prospective cohort study. Crit Care Med 2015; 43:1079–1086
20. Escobar GJ, Laguardia JC, Turk BJ, et al.: Early detection of impending physiologic deterioration among patients who are not in intensive care: Development of predictive models using data from an automated electronic medical record. J Hosp Med 2012; 7:388–395
21. Frost SA, Alexandrou E, Bogdanovski T, et al.: Unplanned admission to intensive care after emergency hospitalisation: risk factors and development of a nomogram for individualising risk. Resuscitation 2009; 80:224–230
22. Hong W, Earnest A, Sultana P, et al.: How accurate are vital signs in predicting clinical outcomes in critically ill emergency department patients. Eur J Emerg Med 2013; 20:27–32
23. Hunziker S, Stevens J, Howell MD: Red Cell Distribution Width and Mortality in Newly Hospitalized Patients. Am J Med 2012; 125:283–291
24. Loekito E, Bailey J, Bellomo R, et al.: Common laboratory tests predict imminent death in ward patients. Resuscitation 2013; 84:280–285

25. Schuetz P, Hausfater P, Amin D, et al.: Biomarkers from distinct biological pathways improve early risk stratification in medical emergency patients: the multinational, prospective, observational TRIAGE study. Crit Care 2015; 19:377
26. Steiner D, Renetseder F, Kutz A, et al.: Performance of the Manchester Triage System in Adult Medical Emergency Patients: A Prospective Cohort Study. J Emerg Med 2016; 50:678–689
27. Sudarshan M, Feldman LS, St. Louis E, et al.: Predictors of mortality and morbidity for acute care surgery patients. J Surg Res 2015; 193:868–873
28. Tam V, Frost SAA, Hillman KMM, et al.: Using administrative data to develop a nomogram for individualising risk of unplanned admission to intensive care. Resuscitation 2008; 79:241–248
29. Tsai JCH, Cheng C-W, Weng S-J, et al.: Comparison of risks factors for unplanned ICU transfer after ED admission in patients with infections and those without infections. ScientificWorldJournal 2014; 2014:102929
30. Tsai JCH, Weng SJ, Huang CY, et al.: Feasibility of using the predisposition, insult/infection, physiological response, and organ dysfunction concept of sepsis to predict the risk of deterioration and unplanned intensive care unit transfer after emergency department admission. J Chinese Med Assoc 2014; 77:133–141
31. Wunderink RG, Diederich ER, Caramez MP, et al.: Rapid response team-triggered procalcitonin measurement predicts infectious intensive care unit transfers. Crit Care Med 2012; 40:2090–2095
32. Churpek MM, Yuen TC, Park SY, et al.: Derivation of a cardiac arrest prediction model using ward vital signs. Crit Care Med 2012; 40:2102–2108
33. Loekito E, Bailey J, Bellomo R, et al.: Common laboratory tests predict imminent medical emergency team calls, intensive care unit admission or death in emergency department patients. Emerg Med Australas 2013; 25:132–139
34. Stelfox HT, Bagshaw SM, Gao S: A retrospective cohort study of age-based differences in the care of hospitalized patients with sudden clinical deterioration. J Crit Care 2015; 30:1025–1031
35. Hackmann G, Chen M, Chipara O, et al.: Toward a two-tier clinical warning system for hospitalized patients. AMIA . Annu Symp proceedings AMIA Symp 2011; 2011:511–9

36. Edelson DP: Using Electronic Health Record Data to Develop and Validate a Prediction Model for Adverse Outcomes on the Wards Matthew. Crit Care Med 2015; 42:841–848
37. Williams B: National Early Warning System (NEWS 2). 2012.
38. Royal College of Physcians: National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party. RCP London 2017;

Tables and Figures

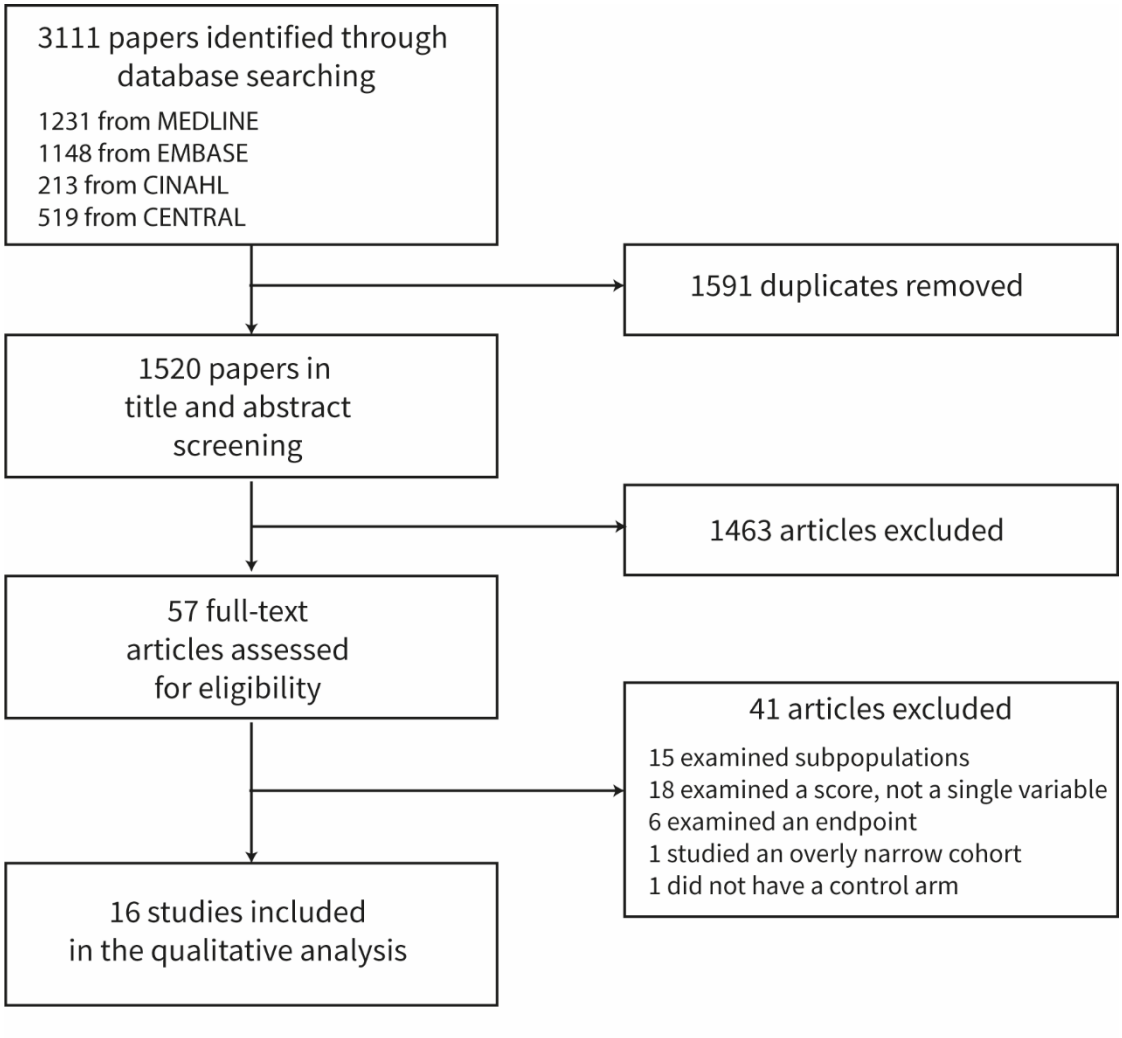


Figure 1. *Flow diagram of the included and excluded studies*

Table 1. *Grading system for strength of evidence*

Strength of Evidence	**Criteria**
Strong	≥2 high-quality studies showing positive association between the presence of a variable and the outcome AND No studies showing a negative association
Moderate	One high-quality AND one lesser-quality study showing association AND No studies showing negative association
Weak	>2 low-quality studies showing positive association OR Only one high-quality study showing positive association
Negative	≥ 1 high-quality study showing negative association (inverse relationship) AND No studies showing a positive association
Inconclusive	Associations present in only one low-quality study OR No studies of any quality showing univariate association OR Presence of positive and negative associations from different articles, regardless of study quality

Table 2. *Details of included studies*

Ref	Lead Author	Publication year	Total number of patients in study	Patients in ICU group	Country	No of sites	Bias scores (high (HQ), medium (MQ) or low quality (LQ))
16	Barfod	2012	6279	102	Denmark	1	5 - MQ
17	Calzavacca	2012	95	15	Australia	1	2 - LQ
18	Churpek	2013	59643	2638	USA	1	7 - HQ
19	Eick	2015	5730	366	Germany	1	4 - LQ
20	Escobar	2012	102,488	3525	USA	14	7 - HQ
21	Frost	2009	126826	1582	Australia	1	7 - HQ
22	Hong	2011	1025	201	Singapore	1	4 - LQ
23	Hunziker	2012	74784	5233	USA	1	5 - MQ
24	Loekiko	2013	70829	149	Australia	2	5 - MQ
25	Schuetz	2015	7000	490	Swiss, France, USA	3	5 - MQ
26	Steiner	2016	2407	93	Switzerland	1	4 - LQ
27	Sudarshan	2015	527	42	USA	1	3 - LQ

28	Tam	2008	94482	672	Australia	1	7 - HQ
29	Tsai	2014	1049	313	Taiwan	1	7 - HQ
30	Tsai	2014	699	214	Taiwan	1	7 - HQ
31	Wunderink	2012	214	71	USA	1	3 - LQ

***Table 3**. Patient centred variables associated with unplanned ICU admission*

Variable	High Quality +'ve Association (ref)	Moderate Quality +'ve Association (ref)	Low Quality +'ve Association (ref)	Negative association (ref)	Overall	Category
History of congestive heart failure (cardiovascular disorder)	21, 29				Strong	Comorbidities
History of diabetes (metabolic disorder)	21, 29				Strong	Comorbidities
Male	20, 21, 28				Strong	Demographic
Increasing age	18, 20, 21, 28	16	27		Strong	Demographic
Diagnosis of hepatic disease (gastrointestinal disorder)	28, 30				Strong	Diagnosis
Higher heart rate (>111 bpm or higher mean in ICU group)	18, 20, 29, 30	16	27		Strong	Vital signs
Higher respiratory rate (>20 bpm or higher in ICU group)	18, 20, 30	16	27		Strong	Vital signs
Higher temperature	18, 20				Strong	Vital signs

Lower arterial oxygen saturation (< 94% or lower in ICU group)	18, 20	16			Strong	Vital signs
Lower diastolic blood pressure	18, 20	16	27		Strong	Vital signs
Lower systolic blood pressure	18, 20				Strong	Vital signs
Female				18, 28	Negative	Demographic
History of respiratory disorder	21		27		Moderate	Comorbidities
Urea (higher in ICU arm)	20	24	27		Moderate	Laboratory tests
White cell count (higher in ICU arm)	20	24			Moderate	Laboratory tests

The number in the boxes are references. Both studies from Tsai et al (29, 30) come from the same patient data base. In accordance with the modified Grading System for Strength of Evidence, these two studies were only counted once (and weighted as a single high-quality study when shown together)

Supplementary Material

SDC-1. Systematic Review Search Design.

SDC-2. Studies excluded because of a single or grouped diagnosis.

SDC-3. Modified Newcastle-Ottawa Scale for assessment of study quality (adapted).

SDC-4. Details (additional) of included studies

SDC-5. Bias scores

SDC-6. Patient populations of included studies

SDC-7. Patient derived variables examined for an association with unplanned ICU admission. In categories and then alphabetical order.

SDC-8. Strength of evidence for individual variables (weak and inconclusive results)

SDC-1. ***Systematic Review Search Design.***

#1. (ICU* or "intensive care" or "critical care").ab,ti.
#2. INTENSIVE CARE UNITS/
#3. CRITICAL CARE/
#4. 1 or 2 or 3
#5. (admission* or admitted or transfer*).ab,ti.
#6. ("risk assessment*" or "risk factor*" or "risk stratif*" or predict* or "increased risk*" or trigger* or score* or scoring or "early warning" or escalat* or deteriorat* or triag* or "vital sign*" or model* or validat*).ab,ti.
#7. 4 and 5 and 6
#8. limit 7 to (humans and yr="2000 Current")
#9. (observational or "case control*" or retrospective or cohort* or "systematic review*").ab,ti.
#10. OBSERVATIONAL STUDY/
#11. CASECONTROL STUDIES/
#12. RETROSPECTIVE STUDIES/
#13. COHORT STUDIES/
#14. RANDOMIZED CONTROLLED TRIAL/
#15. REVIEW/
#16. COMPARATIVE STUDY/
#17. PROSPECTIVE STUDIES/
#18. VALIDATION STUDIES
#19. 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18
#20. 8 and 19
#21. limit 20 to (article or "review")
#22. (unplanned or unexpected or unanticipated or emergency or "rapid response").ti,ab.
#23. 21 and 22

***SDC-2.* Studies excluded because OR a single or grouped diagnosis.**

A. M. Moller,T. Pedersen,N. Villebro,A. Schnaberich,M. Haas,R. Tonnesen. A study of the impact of long-term tobacco smoking on postoperative intensive care admission. Anaesthesia. 2003. 58:55-9
G. Z. Gabayan,M. K. Gould,R. E. Weiss,N. Patel,K. A. Donkor,V. Y. Chiu,S. C. Yiu,J. P. Jones,J. R. Hoffman,C. A. Sarkisian. Poor Outcomes after Emergency Department Discharge of the Elderly: A Case-Control Study. Annals of Emergency Medicine. 2016. 68:43-51
J. C. Knott,S. L. Tan,A. C. Street,M. Bailey,P. Cameron. Febrile adults presenting to the emergency department: Outcomes and markers of serious illness. Emergency Medicine Journal. 2004. 21:170-174
J. Deibener-Kaminsky,J. F. Lesesve,S. Grosset,L. Pruna,M. C. Schmall-Laurain,A. Benetos,P. Kaminsky. [Clinical relevance of leukocyte differential in patients with marked leukocytosis in the emergency room]. Signification d'une hyperleucocytose marquee et de la formule sanguine dans les situations d'urgence.. 2011. 32:406-10
J. M. Caterino,R. A. Murden,K. B. Stevenson. Functional Status Does Not Predict Complicated Clinical Course in Older Adults in the Emergency Department with Infection. Journal of the American Geriatrics Society. 2012. 60:304-309 6p
L. Durairaj,B. Reilly,K. Das,C. Smith,C. Acob,S. Husain,M. Saquib,P. Ganschow,A. Evans,R. McNutt. Emergency department admissions to inpatient cardiac telemetry beds: A prospective cohort study of risk stratification and outcomes. American Journal of Medicine. 2001. 110:7-11
M. Garrouste-Orgeas,A. Boumendil,D. Pateron,P. Aergerter,D. Somme,T. Simon,B. Guidet. Selection of intensive care unit admission criteria for patients aged 80 years and over and compliance of emergency and intensive care unit physicians with the selected criteria: An observational, multicenter, prospective study. Critical Care Medicine. 2009. 37:2919-2928 10p
M. K. Jessen,J. Mackenhauer,A. M. S. W. Hvass,U. Heide-Jorgensen,C. F. Christiansen,H. Kirkegaard. Predictors of intensive care unit transfer or death in emergency department patients with suspected infection. European Journal of Emergency Medicine. 2015. 22:176-180
M. Kennedy,N. Joyce,M. D. Howell,J. L. Mottley,N. I. Shapiro. Identifying Infected Emergency Department Patients Admitted to the Hospital Ward at Risk of Clinical

Deterioration and Intensive Care Unit Transfer. Academic Emergency Medicine. 2010. 17:1080-1085 6p
N. Peschanski,C. Chenevier-Gobeaux,L. Mzabi,R. Lucas,S. Ouahabi,V. Aquilina,V. Brunel,G. Lefevre,P. Ray. Prognostic value of PCT in septic emergency patients. Annals of Intensive Care. 2016.
S. Küpper,C. J. Karvellas,R. G. Khadaroo,S. L. Widder. Increased health services use by severely obese patients undergoing emergency surgery: a retrospective cohort study. Canadian Journal of Surgery. 2015. 58:41-47 7p
T. Olsson,L. Lind. Comparison of the rapid emergency medicine score and APACHE II in nonsurgical emergency department patients. Academic Emergency Medicine. 2003. 10:1040-1048
T. Smith,D. Den Hartog,T. Moerman,P. Patka,E. M. M. Van Lieshout,N. W. L. Schep. Accuracy of an expanded early warning score for patients in general and trauma surgery wards. British Journal of Surgery. 2012. 99:192-197

SDC-3. Modified Newcastle-Ottawa Scale for assessment of study quality (adapted).

Participant Selection	**Score**
Cohort studies	
Selected cohort is *very representative* of the general hospitalised population *(Ward patients scored 2 points)*	**2**
Selected cohort is *somewhat representative* of the general hospitalised population *(ED patients, ward patients recruited via RRS and ward patients with a surgical diagnosis scored 1 point)*	**1**
Cohort is not representative of the general hospital population or the selection of the group was not described *(No such papers are included in this review because they were excluded as per our criteria)*	**0**
Case-control studies	
Cases and controls drawn from the same population and population is *very representative* of the general hospitalised population *(Ward patients scored 2 points)*	**2**
Cases and controls drawn from the same population and population is *somewhat representative* of the general hospitalised population *(ED patients and ward patients recruited via RRS scored 1 point)*	**1**
Cases and controls drawn from different sources or the selection of groups was not described *(No such papers are included in this review because they were excluded as per our criteria)*	**0**
Comparability of groups	
No differences between the groups explicitly reported unless it was one of the variables that was under investigation, or such differences were adjusted for *(The ICU and control arms had to be compared directly and show not to be different to score 2 points)*	**2**
Differences between groups were not recorded *(If the ICU and control arms were not described, the assumption being they may or may not have differed)*	**1**
Groups differed *(If the ICU and control arms were described and shown to be different)*	**0**

Size	
> 100 participants in each group	**2**
< 100 participants in each group	**1**
Confounding	
Adjusted for confounders in study methodology *(Patient selection and recruitment was non biased)*	**2**
Adjusted for confounders during data analysis *(Once selected the ICU and control arm variables underwent univariate analysis with any adjustment described in the methodology)*	**1**
Did not adjust for confounders *(No methodological or analytical description of adjustment)*	**0**

SDC-4. ***Details (additional) of included studies***

Ref	Lead Author	Study type	Prospective or Retrospective	ICU admission primary or secondary endpoint?	Method of variable selection	Cohort
16	Barfod	Cohort	Prospective	Primary	Based on existing scores	Ward patients admitted via ED
17	Calzavacca	Cohort	Prospective	Secondary	Defined by testing device	Ward patients reviewed by RRS team
18	Churpek	Cohort	Retrospective	Primary	Not specified	Ward patients
19	Eick	Cohort	Prospective	Secondary	Scientific hypothesis	ED patients
20	Escobar	Case control	Retrospective	Primary	Derived from previous study by same author	Ward patients
21	Frost	Case control	Retrospective	Primary	Based on existing scores	Ward patients
22	Hong	Cohort	Prospective	Primary	Dervided from Patient Acuity Category traige scale	ED patients

23	Hunziker	Case control	Prospective	Secondary	Scientific hypothesis	Ward patients
24	Loekiko	Cohort	Retrospective	Primary	Statistically selected from a list of 30 lab tests as most predictive of death	ED patients
25	Schuetz	Cohort	Prospective	Secondary	Scientific hypothesis	ED patients
26	Steiner	Cohort	Prospective	Secondary	Based on existing scores	ED patients
27	Sudarshan	Cohort	Retrospective	Primary	Not specified	Emergency surgical admissions ward patients
28	Tam	Cohort	Retrospective	Primary	Not specified	Ward patients
29	Tsai	Case control	Retrospective	Primary	Based on existing scores	Ward patients admitted via ED
30	Tsai	Case control	Retrospective	Primary	Based on existing scores	Ward patients admitted via ED
31	Wunderink	Cohort	Prospective	Primary	Scientific hypothesis	Ward patients reviewed

						by RRS team

SDC-5. ***Bias scores***

Reference	Total	Score of cohort represent-ativeness	Score for group differences	Score for group sizes	Score for adjustment for confounding
16	**5 - MQ**	1	2	2	0
17	**2 - LQ**	1	1	0	0
18	**7 - HQ**	2	2	2	1
19	**4 - LQ**	1	1	2	0
20	**7 - HQ**	2	2	2	1
21	**7 - HQ**	2	2	2	1
22	**4 - LQ**	1	1	2	0
23	**5 - MQ**	2	1	2	0
24	**5 - MQ**	1	1	2	1
25	**5 - MQ**	1	1	2	1
26	**4 - LQ**	1	1	2	0
27	**3 - LQ**	1	1	1	0
28	**7 - HQ**	2	2	2	1
29	**7 - HQ**	2	2	2	1
30	**7 - HQ**	2	2	2	1
31	**3 - LQ**	1	1	1	0

SDC-6. Patient populations of included studies

Ref	ED patient	Ward patient admitted via ED	Ward patient admitted from any source	Ward patient reviewed by RRT/MET
16		*		
17				*
18			*	
19	*			
20			*	
21		*		
22	*			
23			*	
24	*			
25	*			
26	*			
27		*		
28			*	
29		*		
30		*		
31				*

ED: Emergency Departmant

RRT: Rapid Response Team

MET: Medical Emergency Team

SDC-7. Patient derived variables examined for an association with unplanned ICU admission. In categories and then alphabetical order.

Ref	Study quality	Variable	Variable type	ICU vs Control p value	OR	95%CI
28	LQ	Alcohol consumption (history of)	Comorbidities	Non significant		
28	LQ	ASA Class > 3 [lower in ICU arm]	Comorbidities	<0.001		
21	HQ	Cardiovascular disorder, acute MI (history of)	Comorbidities	0.2	1.2	1.0–1.4
28	LQ	Cardiovascular disorder, arrhythmia (history of)	Comorbidities	Non significant		
21	HQ	Cardiovascular disorder, congestive heart failure (history of)	Comorbidities	0.02	1.5	1.3–1.7
30	HQ	Cardiovascular disorder, congestive heart failure (history of)	Comorbidities	0.001	2.2	1.4–3.4
28	LQ	Cardiovascular disorder, coronary artery disease (history of)	Comorbidities	Non significant		
30	HQ	Cardiovascular disorder, coronary artery disease (history of)	Comorbidities	0.02	1.7	1.1–2.5
28	LQ	Cardiovascular disorder, hypertension (history of)	Comorbidities	Non significant		
21	HQ	Cardiovascular disorder, peripheral vascular disease (history of)	Comorbidities	<0.001	1.6	1.3–1.9
21	*HQ*	*Connective tissue disorder (history of)*	*Comorbidities*	*<0.001*	*1.4*	*1.0–2.0*

21	HQ	Gastrointestinal disorder, hepatic disease (severe) (history of)	Comorbidities	0.2	1.3	0.8–2.0
30	HQ	Gastrointestinal disorder, hepatic disease (severe) (history of)	Comorbidities	<0.001	4.4	2.5–7.9
21	*HQ*	*Gastrointestinal disorder, hepatic disease [history of)*	*Comorbidities*	*1*	*1.4*	*1.1–1.9*
30	HQ	Gastrointestinal disorder, hepatic dysfunction (history of)	Comorbidities	0.11	2.6	0.8–7.7
28	LQ	Gastrointestinal disorder, hepatobiliary disease (history of)	Comorbidities	Non significant		
21	HQ	Gastrointestinal disorder, peptic ulcer disease (history of)	Comorbidities	0.05	1	0.8–1.3
21	HQ	Infective disorder, HIV (history of)	Comorbidities	NA		
30	HQ	Infective disorder, immunocompromise (history of)	Comorbidities	0.19	2.1	0.7–6.8
28	LQ	Malignancy (history of)	Comorbidities	Non significant		
21	HQ	Malignancy (history of)	Comorbidities	<0.001	1.4	1.2–1.7
30	HQ	Malignancy, advanced (history of)	Comorbidities	0.54	1.2	0.7–2.2
21	*HQ*	*Malignancy, metastatic (history of)*	*Comorbidities*	*<0.001*	*1.2*	*0.9–1.6*
21	HQ	Metabolic disorder, diabetes (history of)	Comorbidities	0.001	1.3	1.2–1.5

30	HQ	Metabolic disorder, diabetes (history of)	Comorbidities	0.002	1.8	1.3–2.6
28	LQ	Metabolic disorder, diabetes type 2 (history of)	Comorbidities	Non significant		
21	HQ	Metabolic disorder, diabetes with complications (history of)	Comorbidities	<0.001	1.3	1.1–1.6
28	LQ	Metabolic disorder, endocrine disease (history of)	Comorbidities	Non significant		
30	HQ	Neurological disorder, cerebral performance category of 3 or 4 (history of)	Comorbidities	<0.001	3.3	1.9–5.7
28	LQ	Neurological disorder, cerebrovascular disease (history of)	Comorbidities	Non significant		
30	HQ	Neurological disorder, cerebrovascular disease (history of)	Comorbidities	0.08	1.5	1.0–2.3
21	HQ	Neurological disorder, cerebrovascular disease (history of)	Comorbidities	<0.001	1.3	1.1–1.5
28	LQ	Neurological disorder, dementia (history of)	Comorbidities	0.01		
21	HQ	Neurological disorder, dementia (history of)	Comorbidities	0.01	0.8	0.7–0.9
21	*HQ*	*Neurological disorder, paraplegia (history of)*	*Comorbidities*	*0.006*	*1.1*	*0.9–1.4*
21	HQ	*Renal disorder (history of)*	*Comorbidities*	*0.2*	*1.7*	*1.4–2.0*
28	LQ	Renal disorder, chronic kidney disease (history of)	Comorbidities	Non significant		

30	HQ	Renal disorder, end-stage renal disease (history of)	Comorbidities	0.03	2.5	1.2–5.3
21	HQ	Respiratory disorder (history of)	Comorbidities	0.007	1.4	1.2–1.6
28	LQ	Respiratory disorder (history of)	Comorbidities	0.02		
30	HQ	Respiratory disorder, respiratory failure (history of)	Comorbidities	<0.001	7.3	2.8–19.4
28	LQ	Smoking status, Ex-Smoker (history of)	Comorbidities	Non significant		
16	MQ	Age (yrs) [OR per year]	Demographic		1.02	1.01–1.03
30	HQ	Age, ≥65 (yrs) [more in this demographic in ICU group]	Demographic	0.14	1.3	0.9–1.8
21	HQ	Age, 25-34 (yrs) [IRR reference 15-24]	Demographic		1^	0.1-1.3
29	HQ	Age, 30-44 (yrs) [IRR reference 15-29]	Demographic		1.7^	1.2-2.4
21	HQ	Age, 35-44 (yrs) [IRR reference 15-24]	Demographic		1.3^	0.9-1.7
21	HQ	Age, 45-54 (yrs) [IRR reference 15-24]	Demographic		2.2^	1.7-2.9
29	HQ	Age, 45-59 (yrs) [IRR reference 15-29]	Demographic		4.0^	2.9-5.6
21	HQ	Age, 55-64 (yrs) [IRR reference 15-24]	Demographic		2.9^	2.3-3.9
29	HQ	Age, 60-74 (yrs) [IRR reference 15-29]	Demographic		9.0^	6.6-12.2
21	HQ	Age, 65-74 (yrs) [IRR reference 15-24]	Demographic		3.9^	3.1-5.1
21	HQ	Age, 75+ (yrs) [IRR reference 15-24]	Demographic		3.2^	2.5-4.2

29	HQ	Age, 75+ (yrs) [IRR reference 15-29]	Demographic		10.5^	7.7-14.2
21	HQ	Age, mean (yrs) [higher in ICU group]	Demographic	<0.001		
18	HQ	Age, mean (yrs)[higher in ICU group]	Demographic	<0.05		
20	HQ	Age, mean (yrs)[higher in ICU group]	Demographic	<0.001		
28	LQ	Age, median (yrs) [higher in ICU group]	Demographic	0.0046		
29	HQ	Gender, Female (%) [lower in ICU group]	Demographic	<0.001		
18	HQ	Gender, Female (%) [lower in ICU group]	Demographic	<0.05		
28	LQ	Gender, Male (%)[higher in ICU group]	Demographic	Non significant		
20	HQ	Gender, Male (%)[increased in ICU group]	Demographic	<0.001		
21	HQ	Gender, Male [IRR reference female]	Demographic	<0.001	1.38^	1.3-1.5
29	HQ	Gender, Male [IRR reference female]	Demographic		1.7^	1.5-2.0
18	HQ	Race, Black (%)	Demographic	Not significant		
18	HQ	Race, Other (%)	Demographic	Not significant		
18	HQ	Race, Unknown (%)	Demographic	Not significant		
18	HQ	Race, White (%)	Demographic	Not significant		
26	LQ	Cardiovascular disorder (diagnosis of)	Diagnosis		2.9	1.7–4.9

29	HQ	Cardiovascular disorder, acute myocardial infarction (history of) [higher in ICU arm]	Diagnosis	<0.001	2.6 ^	1.8-3.8
29	HQ	Cardiovascular disorder, congestive heart failure (history of)	Diagnosis	<0.001	2.8 ^	1.8-4.3
30	HQ	Cardiovascular disorder, hypotension (diagnosis of)	Diagnosis	<0.001	4	2.0–7.9
31	HQ	Cardiovascular disorder, hypotension (diagnosis of)	Diagnosis		5	2.7–9.3
19	LQ	Deceleration Capacity, mean (m/s) (lower in ICU group)	Diagnosis	<0.001		
29	HQ	Femur (fracture) [higher in ICU arm][OR reference no fracture]	Diagnosis	<0.001	5.7 ^	3.8-8.6
26	LQ	Gastrointestinal disorder (diagnosis of)	Diagnosis		2	0.4–9.5
29	HQ	Gastrointestinal disorder, hepatic disease disease [higher in ICU arm][IRR reference no disease] (diagnosis of)	Diagnosis	<0.001	11.1^	5.9-20.9
31	HQ	Gastrointestinal disorder, hepatic dysfunction (diagnosis of)	Diagnosis		6.3	1.6–23.8
29	HQ	Gastrointestinal disorder, pancreatitis (diagnosis of) [IRR reference no pancreatitis]	Diagnosis		4.1 ^	2.4-7.2
30	HQ	Haematological dysfunction (diagnosis of)	Diagnosis	<0.001	3.7	2.1–6.7

31	HQ	Haematological dysfunction (diagnosis of)	Diagnosis		4.4	2.4–8.1
26	LQ	Infective disorder (diagnosis of)	Diagnosis		2.2	1.1–4.5
26	LQ	Malignancy (diagnosis of)	Diagnosis		3.6	0.6–22.2
26	LQ	Metabolic disorder (diagnosis of)	Diagnosis		1.2	0.1–9.5
30	HQ	Metabolic disorder, dysfunction (diagnosis of)	Diagnosis	<0.001	8.1	2.9–23.1
31	HQ	Metabolic disorder, dysfunction (diagnosis of)	Diagnosis		7.4	2.9–18.9
26	LQ	Neurological disorder (diagnosis of)	Diagnosis		7.8	3.3–18.6
29	HQ	Neurological disorder, cerebral infarction (diagnosis of) [higher in ICU arm][IRR reference no disease]	Diagnosis	<0.001	3.4 ^	1.9-6.1
26	LQ	Other Diagnosis	Diagnosis		3	1.1–7.8
30	HQ	Renal disorder, dysfunction (diagnosis of)	Diagnosis	0.002	2.3	1.4–3.8
31	HQ	Renal disorder, dysfunction (diagnosis of)	Diagnosis		2.9	1.7–4.7
26	LQ	Respiratory disorder (diagnosis of)	Diagnosis		1.7	0.5–5.6
29	HQ	Respiratory disorder, asthma (diagnosis of)[IRR reference no disease]	Diagnosis		0.8 ^	0.3-2.0
29	HQ	Respiratory disorder, COPD (diagnosis of) [higher in ICU arm][IRR reference no disease]	Diagnosis	<0.001	3.6 ^	2.5-4.5

30	HQ	Respiratory disorder, dysfunction (diagnosis of)	Diagnosis	<0.001	4.3	2.3–8.0
31	HQ	Respiratory disorder, dysfunction (diagnosis of)	Diagnosis		3.3	2.0–5.6
29	HQ	Respiratory disorder, pneumonia (diagnosis of) [higher in ICU arm][IRR reference no disease]	Diagnosis	<0.001	4.4 ^	3.3-5.8
24	MQ	Albumin, mean (g/L) [lower in ICU arm][OR threshold ≤ 33.5]	Laboratory tests	<0.0001	5.32	3.88-7.3
24	*MQ*	*Bilirubin, median (umol/L) [higher in ICU arm][OR threshold ≥ 14.5]*	*Laboratory tests*	*0.59*	*1.68*	*1.22-2.33*
17	LQ	BNP, mean (ng/ml)	Laboratory tests	Not reported		
25	MQ	Copeptin, increase in (pmol/L)	Laboratory tests		3	2.6–3.4
28	LQ	Creatinine, median (mmol/L)	Laboratory tests	Non significant		
24	MQ	Creatinine, median (mmol/L) [higher in ICU arm][OR threshold ≥ 105.5]	Laboratory tests	<0.0001	6.04	4.51-8.09
17	LQ	D-dimer, mean (ng/ml)	Laboratory tests	Not reported		
24	MQ	Haematocrit, mean (L/L) [lower in ICU arm][OR threshold ≤ 0.345]	Laboratory tests	0.005	2.3	1.69-3.12
20	HQ	Haematocrit, mean [lower in ICU arm]	Laboratory tests	<0.103		
24	MQ	Haemoglobin, mean (g/L) [lower in ICU arm][OR threshold ≥ 106.5]	Laboratory tests	<0.0001	3.07	2.22-4.24

24	MQ	pH, mean [lower in ICU arm][OR threshold ≤ 7.315]	Laboratory tests	<0.0001	7.68	6.04-9.78
25	MQ	Pro-adrenomedullin, increase in (nmol/L)	Laboratory tests		7.7	5.8–10.3
25	MQ	Procalcitonin, increase in (ug/L)	Laboratory tests		1.9	1.6–2.1
32	LQ	Procalcitonin, median (ng/ml) [higher in ICU arm]	Laboratory tests	<0.0001		
23	MQ	Red cell distribution width, increase in deciles (%)	Laboratory tests		1.15	1.14–1.17
24	MQ	Total bicarbonate, mean (mmol/L) [lower in ICU arm]	Laboratory tests	<0.0001		
20	HQ	Urea, mean (mmol/L) [higher in ICU arm]	Laboratory tests	<0.001		
28	LQ	Urea, median (mmol/L) [higher in ICU arm]	Laboratory tests	0.0219		
24	MQ	Urea, median (mmol/L) [higher in ICU arm][OR threshold ≥ 9.05]	Laboratory tests	<0.0001	3.54	2.64-4.75
28	LQ	White cell count, <12 (10x9/L)	Laboratory tests	Non significant		
28	LQ	White cell count, <4 (10x9/L) [higher % of in ICU arm]	Laboratory tests	<0.001		
30	HQ	White cell count, >12,000 or <4000 (/uL) or band > 5 (%) (diagnosis of)	Laboratory tests	<0.001	3.9	2.1–7.3
31	HQ	White cell count, >12,000 or <4000 (/uL) or band > 5 (%) (diagnosis of)	Laboratory tests		1.7	1.2–2.4

20	HQ	White cell count, mean (10x9/L) [higher in ICU arm]	Laboratory tests	<0.001		
24	*MQ*	*White cell count, median (10x9/L) [higher in ICU arm][OR threshold ≥ 12.05]*	*Laboratory tests*	*<0.47*	*3.18*	*2.38-4.24*
28	LQ	Medication, Anti-Coagulant (patient taking)	Medication	<0.001		
28	LQ	Medication, Anti-Platelet (patient taking)	Medication	Non significant		
28	LQ	Medication, Corticosteriod (patient taking)	Medication	0.076		
28	LQ	Medication, Total count of (median) [higher in ICU arm]	Medication	0.0022		
30	HQ	Cardiovascular disorder, chest pain (new onset) (symptom/signs of)	Symptom/ Sign	0.003	4.4	1.6–11.7
26	LQ	Infective disorder, fever (complaint of)	Symptom/ Sign		1.6	0.5–5.1
26	LQ	Neurological disorder (symptom of)	Symptom/ Sign		7.5	3.2–17.5
30	HQ	Neurological disorder, altered mental status (symptom/signs of)	Symptom/ Sign	<0.001	31	7.1–134.6
30	HQ	Neurological disorder, seizure (symptom/signs of)	Symptom/ Sign	0.02	3	1.2–7.4
26	LQ	Other Presenting Complaint	Symptom/ Sign		6.2	1.7–23.0
26	LQ	Pain, except chest pain (symptom of)	Symptom/ Sign		2.7	1.5–5.0

26	LQ	Respiratory disorder (symptom of)	Symptom/ Sign		1.6	0.9–3.0
30	HQ	Respiratory disorder (symptoms/signs of)	Symptom/ Sign	<0.001	4.5	2.4–8.4
26	LQ	Thoracic pain (symptom of)	Symptom/ Sign		3.3	1.4–7.8
16	MQ	Arterial oxygen saturation, <80 (%) [OR reference 95-100]	Vital signs		8.42	1.88–36.17
22	LQ	Arterial oxygen saturation, <95 (%) [OR reference ≥ 95]	Vital signs		1.32	0.88-1.97
16	MQ	Arterial oxygen saturation, 80-89 (%) [OR reference 95-100]	Vital signs		7.49	3.86–14.51
16	MQ	Arterial oxygen saturation, 90-94 (%) [OR reference 95-100]	Vital signs		2.35	1.31–4.20
20	HQ	Arterial oxygen saturation, mean (%) [lower in ICU group]	Vital signs	<0.001		
18	HQ	Arterial oxygen saturation, median [lower in ICU group]	Vital signs	<0.05		
22	LQ	Blood pressure, diastolic, <60 (mmHg) [OR reference 60-95]	Vital signs		1	0.67-1.49
22	LQ	Blood pressure, diastolic, >95 (mmHg) [OR reference 60-95]	Vital signs		1.13	0.76-1.70
18	HQ	Blood pressure, diastolic, mean (mmHg) [lower in ICU group]	Vital signs	<0.05		

20	HQ	Blood pressure, diastolic, mean (mmHg) [lower in ICU group]	Vital signs	<0.001		
16	MQ	Blood pressure, systolic, <80 (mmHg) [OR reference 90-]	Vital signs		3.22	0.43–24.32
28	LQ	Blood pressure, systolic, <90 (mmHg) [lower in ICU group]	Vital signs	0.002		
22	LQ	Blood pressure, systolic, <90 (mmHg) [OR reference 90-140]	Vital signs		0.95	0.57-1.56
22	LQ	Blood pressure, systolic, >140 (mmHg) [OR reference 90-140]	Vital signs		0.74	0.53-1.03
16	MQ	Blood pressure, systolic, 80-89 (mmHg) [OR reference 90-]	Vital signs		4.97	1.50–16.38
18	HQ	Blood pressure, systolic, mean (mmHg) [lower in ICU group]	Vital signs	<0.05		
20	HQ	Blood pressure, systolic, mean (mmHg) [lower in ICU group]	Vital signs	<0.001		
22	LQ	Glasgow Coma Scale, <15 (out of 15) [OR reference 15/15]	Vital signs		2.41	1.63-3.55
16	MQ	Glasgow Coma Scale, <8 (out of 15) [OR reference 15/15]	Vital signs		5.21	5.24–24.91
16	MQ	Glasgow Coma Scale, 14 (out of 15) [OR reference 15/15]	Vital signs		3.57	1.82–7.00

16	MQ	Glasgow Coma Scale, 9-13 (out of 15) [OR reference 15/15]	Vital signs		2.02	0.73–5.61
16	MQ	Heart rate, <40 (bpm) [OR reference 50-110]	Vital signs		6.51	0.84–50.50
22	LQ	Heart rate, <60 (bpm) [OR reference 60-100]	Vital signs		2.25	1.32-3.81
28	LQ	Heart rate, >100 (bpm) [lower in ICU group]	Vital signs	0.007		
22	LQ	Heart rate, >100 (bpm) [OR reference 60-100]	Vital signs		0.95	0.68-1.32
16	MQ	Heart rate, >130 (bpm) [OR reference 50-110]	Vital signs		8.14	4.07–16.28
30	HQ	Heart rate, ≥130 (bpm)	Vital signs	0.03	1.6	1.1–2.3
31	HQ	Heart rate, ≥130 (bpm)	Vital signs		3.7	2.1–6.4
16	MQ	Heart rate, 111-120 (bpm) [OR reference 50-110]	Vital signs		3.83	2.05–7.18
16	MQ	Heart rate, 121-130 (bpm) [OR reference 50-110]	Vital signs		8.91	4.68–16.95
16	MQ	Heart rate, 40-49 (bpm) [OR reference 50-110]	Vital signs		1.41	0.20–10.33
18	HQ	Heart rate, mean (bpm) [higher in ICU group]	Vital signs	<0.05		
20	HQ	Heart rate, mean (bpm) [higher in ICU group]	Vital signs	<0.001		
18	HQ	Mental status, Alert [AVPU scale]	Vital signs	Non significant		
18	HQ	Mental status, Responsive to pain [AVPU scale]	Vital signs	Non significant		
18	HQ	Mental status, Responsive to voice [AVPU scale]	Vital signs	Non significant		

18	HQ	Mental status, Unresponsive [AVPU scale]	Vital signs	Non significant		
28	LQ	Respiratory rate, >20 (bpm) (% above or below 20 bpm)[lower in ICU group]	Vital signs	<0.001		
22	LQ	Respiratory rate, >20 (bpm) [OR reference 12-20]	Vital signs		1.06	0.70-1.60
16	MQ	Respiratory rate, >35 (bpm) [OR reference 8-25]	Vital signs		9.11	3.49–23.80
31	HQ	Respiratory rate, ≥30 (bpm)	Vital signs		3.2	1.7–6.0
16	MQ	Respiratory rate, 26-30 (bpm) [OR reference 8-25]	Vital signs		3.76	1.97–7,20
16	MQ	Respiratory rate, 31-35 (bpm) [OR reference 8-25]	Vital signs		5.92	2.08–16.87
18	HQ	Respiratory rate, mean (bpm) [higher in ICU group]	Vital signs	<0.05		
20	HQ	Respiratory rate, mean (bpm) [higher in ICU group]	Vital signs	<0.001		
31	HQ	Temperature > 38 or < 36 (degress C)	Vital signs		0.8	0.6-1.2
28	LQ	Temperature, <36 (degrees C)	Vital signs	Non significant		
28	LQ	Temperature, >38 (degrees C)	Vital signs	Non significant		
18	HQ	Temperature, mean (degrees C) [higher in ICU group]	Vital signs	<0.05		
20	HQ	Temperature, mean (degrees F) [higher in ICU group]	Vital signs	0.009		

16	MQ	Vital sign count (1 abnormal vital sign) [OR reference 0 abnormal vital signs]	Vital signs		2.2	1.25–3.89
16	MQ	Vital sign count (2 abnormal vital signs) [OR reference 0 abnormal vital signs]	Vital signs		13.03	7.64–22.23
16	MQ	Vital sign count (3 abnormal vital signs) [OR reference 0 abnormal vital signs]	Vital signs		15.99	6.76–37.77

If only one of p value and RRs/ORs/IRRs were significant, the variable was considered significant.

Black denotes a significant p value < 0.05 or Odds Ratio, Incidence Rate Ratio, Risk Ratio

Grey denotes a non-significant p value < 0.05 or Odds Ratio, Incidence Rata Ratio, Risk Ratio

Italics denotes a difference in significance of p value < 0.05 and OR/RR/IRR

^Denotes the value is an Incidence rate ratio, Risk Ratio [29]

SDC-8. ***Strength of evidence for individual variables (weak and inconclusive results)***

Variable	High Quality +'ve Association	Moderate Quality +'ve Association	Low Quality +'ve Association	Negative	Overall	Category
Cardiovascular disorder, coronary artery disease (history of)	29				Weak	Comorbidities
Cardiovascular disorder, peripheral vascular disease (history of)	21				Weak	Comorbidities
Connective tissue disorder (history of)	21				Weak	Comorbidities
Gastrointestinal disorder, hepatic disease (severe) (history of)	29				Weak	Comorbidities
Gastrointestinal disorder, hepatic disease [history of)	21				weak	Comorbidities
Malignancy (history of)	21				Weak	Comorbidities
Malignancy, metastatic (history of)	21				Weak	Comorbidities

Metabolic disorder, diabetes with complications (history of)	21				Weak	Comorbidities
Neurological disorder, cerebral performance category of 3 or 4 (history of)	29				Weak	Comorbidities
Neurological disorder, cerebrovascular disease (history of)	21				Weak	Comorbidities
Neurological disorder, paraplegia (history of)	21				Weak	Comorbidities
Renal disorder (history of)	21				Weak	Comorbidities
Renal disorder, end-stage renal disease (history of)	29				Weak	Comorbidities
Respiratory disorder, respiratory failure (history of)	29				Weak	Comorbidities
Cardiovascular disorder, acute	28				Weak	Diagnosis

myocardial infarction (history of) [higher in ICU arm]						
Cardiovascular disorder, congestive heart failure (history of)	28				Weak	Diagnosis
Cardiovascular disorder, hypotension (diagnosis of)	29, 30				Weak	Diagnosis
Femur (fracture) [higher in ICU arm][OR reference no fracture]	28				Weak	Diagnosis
Gastrointestinal disorder, pancreatitis (diagnosis of) [IRR reference no pancreatitis]	28				Weak	Diagnosis
Haematological dysfunction (diagnosis of)	29, 30				Weak	Diagnosis
Metabolic disorder, dysfunction (diagnosis of)	29, 30				Weak	Diagnosis

Neurological disorder, cerebral infarction (diagnosis of) [higher in ICU arm][IRR reference no disease]	28				Weak	Diagnosis
Renal disorder, dysfunction (diagnosis of)	29, 30				Weak	Diagnosis
Respiratory disorder, COPD (diagnosis of) [higher in ICU arm][IRR reference no disease]	28				Weak	Diagnosis
Respiratory disorder, dysfunction (diagnosis of)	29, 30				Weak	Diagnosis
Respiratory disorder, pneumonia (diagnosis of) [higher in ICU arm][IRR reference no disease]	28				Weak	Diagnosis
Procalcitonin, median (ng/ml)		25	31		Weak	Laboratory tests

[higher in ICU arm]						
White cell count, >12,000 or <4000 (/uL) or band > 5 (%)	29, 30				Weak	Laboratory tests
Cardiovascular disorder, chest pain (new onset) (symptom/signs of)	29				Weak	Symptom/Sign
Neurological disorder, altered mental status (symptom/signs of)	29				Weak	Symptom/Sign
Neurological disorder, seizure (symptom/signs of)	29				Weak	Symptom/Sign
Respiratory disorder (symptoms/signs of)	29				Weak	Symptom/Sign
Glasgow Coma Scale, <15 (out of 15) [OR reference 15/15]		16	22		Weak	Vital signs
Alcohol abuse (history of)					Inconclusive	Comorbidities
Alcohol abuse (history of)					Inconclusive	Comorbidities

Alcohol consumption (history of)					Inconclusive	Comorbidities
Cardiovascular disorder, acute MI (history of)					Inconclusive	Comorbidities
Cardiovascular disorder, arrhythmia (history of)					Inconclusive	Comorbidities
Cardiovascular disorder, coronary artery disease (history of)					Inconclusive	Comorbidities
Cardiovascular disorder, hypertension (history of)					Inconclusive	Comorbidities
Gastrointestinal disorder, hepatic disease (severe) (history of)					Inconclusive	Comorbidities
Gastrointestinal disorder, hepatic dysfunction (history of)					Inconclusive	Comorbidities
Gastrointestinal disorder, hepatobiliary disease (history of)					Inconclusive	Comorbidities

Gastrointestinal disorder, peptic ulcer disease (history of)					Inconclusive	Comorbidities
Infective disorder, HIV (history of)					Inconclusive	Comorbidities
Infective disorder, immunocompromise (history of)					Inconclusive	Comorbidities
Infective disorder, immunocompromise (history of)					Inconclusive	Comorbidities
Malignancy (history of)					Inconclusive	Comorbidities
Malignancy, advanced (history of)					Inconclusive	Comorbidities
Metabolic disorder, diabetes type 2 (history of)					Inconclusive	Comorbidities
Metabolic disorder, endocrine disease (history of)					Inconclusive	Comorbidities
Neurological disorder, cerebrovascular					Inconclusive	Comorbidities

disease (history of)						
Neurological disorder, cerebrovascular disease (history of)					Inconclusive	Comorbidities
Neurological disorder, cerebrovascular disease (history of)					Inconclusive	Comorbidities
Neurological disorder, dementia (history of)			27	21	Inconclusive	Comorbidities
Renal disorder, chronic kidney disease (history of)					Inconclusive	Comorbidities
Smoking status, Ex-Smoker (history of)					Inconclusive	Comorbidities
Age ≥ 80 years					Inconclusive	Demographic
Cardiovascular disorder (diagnosis of)					Inconclusive	Diagnosis
Cardiovascular disorder (diagnosis of)				26	Inconclusive	Diagnosis

Gastrointestinal disorder (diagnosis of)					Inconclusive	Diagnosis
Infective disorder (diagnosis of)			26		Inconclusive	Diagnosis
Malignancy (diagnosis of)					Inconclusive	Diagnosis
Metabolic disorder (diagnosis of)					Inconclusive	Diagnosis
Neurological disorder (diagnosis of)			26		Inconclusive	Diagnosis
Other Diagnosis			26		Inconclusive	Diagnosis
Respiratory disorder (diagnosis of)					Inconclusive	Diagnosis
Respiratory disorder, asthma (diagnosis of)[IRR reference no disease]					Inconclusive	Diagnosis
Albumin, mean (g/L) [lower in ICU arm][OR threshold ≤ 33.5]		24			Inconclusive	Laboratory tests
Bilirubin, median (umol/L) [higher in ICU		24			Inconclusive	Laboratory tests

arm][OR threshold ≥ 14.5]						
BNP, mean (ng/ml)					Inconclusive	Laboratory tests
Copeptin, increase in (pmol/L)		25			Inconclusive	Laboratory tests
Creatinine, median (mmol/L) [higher in ICU arm][OR threshold ≥ 105.5]		24			Inconclusive	Laboratory tests
D-dimer, mean (ng/ml)					Inconclusive	Laboratory tests
Haematocrit, mean (L/L) [lower in ICU arm][OR threshold ≤ 0.345]		24			Inconclusive	Laboratory tests
Haemoglobin, mean (g/L) [lower in ICU arm][OR threshold ≥ 106.5]		24			Inconclusive	Laboratory tests
pH, mean [lower in ICU arm][OR threshold ≤ 7.305]		24			Inconclusive	Laboratory tests

Pro-adrenomedullin, increase in (nmol/L)		25			Inconclusive	Laboratory tests
Red cell distribution width, increase in deciles (%)		23			Inconclusive	Laboratory tests
Total bicarbonate, mean (mmol/L) [lower in ICU arm]		24			Inconclusive	Laboratory tests
White cell count, < 12 (10x9/L) [higher % of in ICU arm]					Inconclusive	Laboratory tests
White cell count, <4 (10x9/L) [higher % of in ICU arm]			27		Inconclusive	Laboratory tests
Medication, Anti-Coagulant (patient taking)			27		Inconclusive	Medication
Medication, Anti-Platelet (patient taking)					Inconclusive	Medication
Medication, Corticosteriod (patient taking)					Inconclusive	Medication
Medication, Total count of			27		Inconclusive	Medication

(median) [higher in ICU arm]						
Deceleration Capacity, mean (m/s) (lower in ICU group)			19		Inconclusive	N/A
Infective disorder, fever (complaint of)					Inconclusive	Symptom/Sign
Neurological disorder (symptom of)			26		Inconclusive	Symptom/Sign
Other Presenting Complaint			26		Inconclusive	Symptom/Sign
Pain, except chest pain (symptom of)			26		Inconclusive	Symptom/Sign
Respiratory disorder (reason for RRS activation, symtom/signs of)					Inconclusive	Symptom/Sign
Respiratory disorder (symptom of)					Inconclusive	Symptom/Sign
Thoracic pain (symptom of)			26		Inconclusive	Symptom/Sign
Heart rate [<60 (bpm)]			22		Inconclusive	Vital signs

^Both studies from Tsai et al (29)(30) come from the same patient data base. In accordance with the modified Grading System for Strength of Evidence, these two studies were only counted once (and weighted as a single high-quality study when shown together)

Chapter 2. Evaluating the prognostic value of fractional inspired oxygen for potential inclusion in HAVEN

Most Early Warning Score (EWS) systems use peripheral blood oxygen saturation (SpO_2) measurements recorded by a pulse oximeter as one of the vital signs. However, patients with a low SpO_2 are treated by increasing fractional inspired oxygen concentration (FiO_2). This therapy returns their SpO_2 value to normal (higher) values, making FiO_2 an important and likely correlated marker of probable deterioration. Despite this, FiO_2 is not part of any widely implemented EWS, nor is it part of comparable machine learning prognostic risk scores. The National Early Warning Score (NEWS) is the most widely implemented in the UK and internationally and scores oxygen requirements in a binary manner (scoring zero for room air and two for all forms of supplementary oxygen). As a result, important information about respiratory dysfunction in ward patients is potentially being lost. This approach also means that escalations in oxygen therapy occur without an increase in score, creating a risk these changes may be missed by Rapid Response Systems (RRS) reliant on the NEWS. To evaluate the potential prognostic value of FiO_2, an observational study was conducted to test the hypothesis that adding FiO_2 to the NEWS would improve the predictive performance of the score when used in patients requiring oxygen. The results of the study showed adding FiO_2 to NEWS improved performance. Further, the approach was shown to be feasible for use in the HAVEN model.

THE EFFECT OF FRACTIONAL INSPIRED OXYGEN CONCENTRATION ON EARLY WARNING SCORE PERFORMANCE: A DATABASE ANALYSIS

Authors

James Malycha MBBD FCICM[1,2,3], Nazli Farajidavar BSc MSc PhD[4], Marco A. F. Pimentel DPhil[1,7], Oliver Redfern MBBS PhD[1], Lionel Tarassenko CBE FREng FMedSci MA DPhil CEng FIET[7], Paul Meredith PhD[5], David Prytherch PhD CSci MIPEM[6], David A. Clifton MEng DPhil (Oxon) CEng[7], Guy Ludbrook PhD MBBS FANZCA[2], Duncan Young BM DM FRCA FMedSci[1], Peter Watkinson MD ChB MRCP FDICM[1]

Affiliations

[1]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[2]Discipline of Acute Care Medicine, University of Adelaide, South Australia, Australia

[3]Intensive Care Unit, The Queen Elizabeth Hospital, Adelaide, South Australia, Australia

[4]James Black Centre, King's College, London, UK

[5]Research and Innovation Department, Portsmouth Hospitals NHS Trust, Portsmouth, UK

[6]Centre for Healthcare Modelling & Informatics, School of Computing, University of Portsmouth, UK

[7]Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK

Keywords

Critical Care, Intensive Care, Early Warning Scores, Fractional Inspired Oxygen, Predictive Scores, Machine Learning, Decision Trees

Funding

This publication presents independent research funded by the Health Innovation Challenge Fund (HICF-R9-524; WT-103703/Z/14/Z), a parallel funding partnership between the Department of Health and Wellcome Trust.

Statement of Authorship

Title of paper	The Effect of Fractional Inspired Oxygen Concentration on Early Warning Score Performance: a database analysis
Publication status	Published
Publication details	Malycha J*, Farajidavar N*, Pimentel MAF, et al.: The effect of fractional inspired oxygen concentration on early warning score performance: A database analysis. Resuscitation 2019; 139:192–199 *Joint first authors

Principal Author

Name of Principal Author	Dr James Malycha
Contribution to paper	Conceptualised the work, wrote manuscript and was corresponding author
Overall percentage (%)	45% (Joint first author)
Certification	This paper reports on original research conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Co-author contributions

Each author certifies:

- The candidate's stated contribution to the publication is accurate (as state above);
- Permission is granted for the candidate to include the publication in the thesis; and
- The sum of all co-author contributions is equal less the candidate's stated contribution

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Nazli Farajidavar
Contribution to paper	Joint first author, conceptualised the work, managed all data analysis and machine learning elements of the project

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Marco Pimentel
Contribution to paper	Assisted with data analysis, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Oliver Redfern
Contribution to paper	Assisted with data analysis, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor David Clifton
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Lionel Tarassenko
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Paul Meredith
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor David Prytherch
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Guy Ludbrook
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Duncan Young
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Peter Watkinson
Contribution to paper	Evaluated and edited manuscript

Abstract

Objectives

To calculate fractional inspired oxygen concentration (FiO_2) thresholds in ward patients and add these to the National Early Warning Score (NEWS). To evaluate the performance of NEWS-FiO_2 against NEWS when predicting in-hospital death and unplanned intensive care unit (ICU) admission.

Methods

A multi-centre, retrospective, observational cohort study was carried out in five hospitals from two UK NHS Trusts. Adult admissions with at least one complete set of vital sign observations recorded electronically were eligible. The primary outcome measure was an 'adverse event' which comprised either in-hospital death or unplanned ICU admission. Discrimination was assessed using the Area Under the Receiver Operating Characteristic curve (AUROC).

Results

A cohort of 83304 patients from a total of 271363 adult admissions were prescribed oxygen. In this cohort, NEWS-FiO_2 (AUROC 0.823, 95% CI 0.819-0.824) outperformed NEWS (AUORC 0.811, 95% CI 0.809-0.814) when predicting in-hospital death or unplanned ICU admission within 24 hours of a complete set of vital sign observations.

Conclusions

NEWS-FiO_2 generates a performance gain over NEWS when studied in ward patients requiring oxygen. This warrants further study, particularly in patients with respiratory disorders.

Introduction

An Early Warning Score (EWS) identifies clinical deterioration in hospitalised patients using simple algorithms that sum integer scores assigned to values of individual vital sign observations.(1) The score increases as the vital signs become more abnormal. The summed scores are then calibrated against subsequent in-hospital adverse events to generate thresholds to trigger escalations in care.(2) EWS systems were originally designed to be paper-based, but as some vital sign recording has become electronic, more sophisticated systems have been developed.(3)(4)

Most EWS systems use peripheral blood oxygen saturation (SpO_2) recorded by a pulse oximeter as one of the vital signs. However, patients with a low SpO_2 are often treated by increasing their inspired oxygen fractional concentration (FiO_2), which returns their SpO_2 value to normal and makes the FiO_2 the important value for detecting deterioration.(5,6) Despite this, FiO_2 is not part of any widely implemented EWS.(7) Techniques that minimise this information loss by using the fractional inspired oxygen as an alternative or adjunct to SpO_2 to construct an EWS are relatively under-developed, and are mainly used in obstetric populations.(8)(9) The National Early Warning Score (NEWS) is the most widely adopted in the UK and scores in a binary manner for oxygen use (scoring zero for room air and two for all forms of supplementary oxygen).(10)(11) As a result, important information about respiratory dysfunction in ward patients may being lost in NEWS, impairing its performance. The lack of granularity also means that escalations in oxygen therapy may occur without an increase in score, creating a risk of these changes being missed by reviewers. This study was designed to test the hypothesis that adding FiO_2 to NEWS would improve the predictive performance of the score when used in patients requiring oxygen.(12)

Methods

This multi-centre cohort study is reported following the TRIPOD guidelines for the development and validation of predictive models.(13) The TRIPOD checklist is included in the supplementary digital content **(SDC-1)**. We performed a two-centre, retrospective study using two large databases of routinely collected healthcare data. This study was part of a larger project for which ethics approval had previously been obtained (Health Research Authority reference: Oxford University Hospitals Trust Research Ethics Committee reference

16/SC/0264; confidentiality advisory group: 16/CAG/0066, Isle of Wight, Portsmouth and South East Hampshire Research Ethics Committee reference: 08/02/1394).

Source of data

One database contained vital signs, oxygen administration data and patient outcome data on all patients admitted to the four acute care hospitals in the Oxford University Hospitals NHS Foundation Trust (OUHNHSFT) between October 2014 and October 2016 where vital sign recordings were taken at the bedside using the System for Electronic Notification and Documentations (SEND, Sensyne Health, www.sensynehealth.com).(14) The second database contained similar data on patients admitted to the Queen Alexandra Hospital (QAH), an acute care hospital in Portsmouth, between January 2010 and May 2016 where vital sign recordings were taken using CareFlow Vitals (System C Healthcare, www.systemc.com).

Participants

All admissions to the OUHNHSFT hospitals and the QAH were eligible for the study. To be included in the analysis, patients were required to be adult (≥ 16 years of age) at hospital admission, with a hospital stay of ≥ 24 hours with at least one complete vital sign observation set. Vital sign observations sets were only eligible for analysis once the patient reached the ward and having not arrived there via ICU **(Figure 1)**. Each new patient admission was taken as an individual entity as a source of data, meaning vital sign observation sets taken from one patient on subsequent admissions were eligible for inclusion in the analysis.

Outcomes

We used a binary composite 'adverse event' outcome, which comprised in-hospital death or unplanned intensive care unit (ICU) admissions. Where patients were admitted to ICU and subsequently died, the ICU admission was taken as the event.

Predictors

For each vital sign observation set we collected heart rate (HR), respiratory rate (RR), systolic blood pressure (SBP), peripheral blood oxygen saturation (SpO_2), body temperature (Temp), neurological status using the Alert-Verbal-Painful-Unconscious (AVPU) scale, and the

composition (air/oxygen) and delivery method (mask type) of inhaled gas. Where consciousness level had been recorded using the Glasgow Coma Scale system we converted to AVPU using a scoring system shown in the supplementary digital content **(SDC-2)**. The databases also contained survival status at hospital discharge and details of unplanned ICU admissions via linkage to electronic ICU records. We did not analyse vital sign observation sets in patients who were post-ICU admission on the general wards. In both databases the first adverse event identified was the one used in the analysis. Any complete vital sign observation sets in the 24 hours preceding an event were classified as associated with an event.

FiO_2 calculation

FiO_2 was taken as the prescribed value for fixed performance masks. For all other oxygen delivery systems, FiO_2 was calculated using a published formula: FiO_2 = (O_2 Flow Rate + 0.21(Minute volume – O_2 Flow Rate))/Minute volume.(15) We assumed a fixed tidal volume of 0.45 litre per breath for all patients and multiplied this by the respiratory rate to obtain minute volume. A summary of the mask types and corresponding flow rates used in our calculations are shown in the supplementary digital content **(SDC-3)**. We calculated error rates associated with the assumption of a fixed tidal volume (supplementary digital content, **SDC-4**). We assumed a maximum FiO_2 of 1.0 for all patients requiring high flow nasal oxygen and non-invasive ventilation methods.

FiO_2 threshold development using Decision Tree Analysis

A Decision Tree (DT) is a predictive model that can be applied to any numeric or categorical database to establish which variables are most strongly associated with pre-specified outcomes. We adopted the methodology of Badriyah et al. to generate thresholds for the calculated FiO_2.(16) We used the Scikit-learn package within Python 2.7 to carry out our analysis. We generated the FiO_2 thresholds using the OUHNHSFT database.(17) In keeping with NEWS, we assigned weights of zero, one, two and three for FiO_2 as the concentration increased.(3)(11) The FiO_2 thresholds, as well as further detail on the DT analysis, are shown in the supplementary digital content **(SDC-5, 6 and 7)**.

Missing Data

To be included in our two databases, vital sign observation sets needed to be complete with a measurement of each vital sign and the inhaled gas composition/delivery method. **Figure 1** shows the number of excluded vital sign observation sets from the analysis.

Development databases

The OUHNHSFT database was used to derive the FiO_2 threshold scoring bands. The QAH database was used to externally validate the NEWS-FiO_2 score.

Evaluating NEWS and NEWS-FiO_2

Evaluation of NEWS and NEWS-FiO_2 was undertaken in two stages. Firstly, we undertook the analysis on observation sets where oxygen was recorded as having been used during the admission. Secondly, we analysed score performance in all observation sets regardless of oxygen use. The primary performance measure was Area Under the Receiver Operating Characteristic curve (AUROC), which provided an overall measure of model discrimination. (7) AUROC results were reported with 95% confidence intervals, computed via bootstrapping the QAH data (through random sampling while preserving the event class prevalence of approximately 1% and repeating the test 1000 times). We used AUROC to test the ability of NEWS and NEWS-FiO_2 to predict an adverse event up to 24 hours prior. We tested the variation in AUROC for each EWS as the time-to-event window reduced from 24 hours to zero. This evaluation metric showed the change in AUROC performance as the patient neared an adverse event and allowed a comparison of performance across this time frame. We used positive predictive value (PPV) vs sensitivity (also known as precision-recall) curves to show the performance of the scores rather than receiver operating characteristic curves since they de-emphasise the much greater numbers of patients without an adverse event correctly identified as true negatives. We used efficiency curves to show the number of triggers generated at different values for each score as an indication of potential workload implications on the ward. Overall sensitivity, specificity and positive predictive values were also calculated using the suggested thresholds of five or above and seven or above. It was not possible to assess calibration since NEWS does not provide estimates of absolute risk.

Results

A flowchart of study participants is included in the supplementary digital content **Figure 1.**

In the OUHNHSFT Training database there were 71735 eligible admissions. Of those excluded, 28750 were discharged alive in <24 hours, 202 had only incomplete vital sign sets and 19 had events but no observation sets taken <24 hours prior. A total of 42764 admissions (29931 patients) were included for analysis (the difference accounted for by multiple admissions for some patients). Of these, 17012 admissions required oxygen therapy (14028 patients), generating 222156 vital sign observation sets. A total of 6469 vital sign observation sets occurred within the 24 hours prior to an adverse event.

In the QAH Test database there were 250815 eligible admissions. Of those excluded, 31920 were discharged alive <24 hours and 1532 had incomplete vital sign observation sets (there is a lower proportion of patients staying <24 hours in the QAH database because of an ambulatory, short stay care facility within the OUHNHSFT) and zero had events but no observation sets taken <24 hours prior. A total of 217363 admissions (120017 patients) were included for analysis. Of these, 83309 admissions (57467 patients) required oxygen, generating 1055423 vital sign observation sets. A total of 30356 vital sign observation sets were tagged as associated with an adverse event.

Demographics and vital sign observation set characteristics, in both the total and oxygen requiring cohorts, are summarised in **Table 1**.

Table 2 shows the distributions of calculated FiO_2 for all vital sign observation sets in the QAH Test database for the oxygen requiring cohort and categorises them into scoring thresholds. 3883 vital sign observation sets had a calculated FiO_2 value between 21 and 22%. This occurred in patients on very low oxygen flow rates, in conjunction with higher respiratory rates (thus diluting the administered oxygen). The decision tree analysis evaluated the patients linked to these vital sign observation sets as having an equivalent risk as those not receiving oxygen, thus they scored zero points. 234504 vital sign observation sets scored one point, 564712 scored two points (equivalent to the score attributed in NEWS for a patient on any amount of oxygen) and 252090 scored three points. Overall, 46.5% of the vital sign observation sets scored zero, one or three points and 53.5% scored two points.

Figure 2 shows the distribution of FiO_2 concentrations in both the OUHNHSFT and QAH. We report all inspired oxygen concentrations as percentages. We report the distribution of calculated FiO_2 values (for each database) as a percentage of the total cohort of vital sign observation sets. FiO_2 concentrations less than 25% were not shown in the figure because its disproportionate height made it difficult to represent graphically with the other groups. The most common FiO_2 in both databases was 45%, each accounting for >5% of the total vital sign observation sets. The higher percentage of vital sign observation sets with a FiO_2 of 100% in the OUHNHSFT is accounted for by the higher provision of non-invasive and nasal-high-flow cannula ventilation strategies on the wards of this Trust.

Performance of the early warning scores

Table 3 shows the observation level AUROC (with 95% CI) for each scoring system against an outcome of 'adverse event' in the subsequent 24 hours. It also shows the sensitivity, specificity, and positive predictive value of the scores for thresholds of ≥ 5 and ≥ 7 respectively. In the oxygen requiring cohort, NEWS-FiO_2 (AUROC of 0.823, 95% CI 0.819-0.824) out performed NEWS (AUROC of 0.811, 95% CI 0.809-0.814) when predicting in-hospital death or unplanned ICU admission within 24 hours of the observation set. NEWS-FiO_2 also outperformed NEWS in sensitivity, specificity and positive predictive value when using five and seven as trigger thresholds. In terms of admission level performance, NEWS-FiO_2 identified an additional 173 admissions (out of the total 83304 in the oxygen therapy cohort) who went on the have an adverse event.

Figure 3 shows performance metric curves in the oxygen cohort for discriminating vital signs observations followed by in-hospital death or unplanned ICU admission 24 hours preceding the event for the QAH dataset**. Figure 3a** displays the AUROC curves for each score in both the oxygen requiring cohort and the total cohort. **Figure 3b** shows improving performance in both EWS in AUROC over time as observation sets approach the adverse event. The FiO_2 enhanced score outperforms the non-enhanced score throughout and particularly in the oxygen requiring cohort. Efficiency curves are displayed in **Figure 3c** but do not show any obvious performance gains, potentially as a result of the small fraction of true positive results within the entire database. **Figure 3d** shows the precision-recall curves.

Discussion

Statement of key findings

In ward patients requiring oxygen therapy, NEWS-FiO_2 outperformed NEWS when predicting in-hospital death or unplanned ICU admission within 24 hours. Our results support the hypothesis that introducing FiO_2 thresholds to increase the granularity of oxygen therapy scores from a binary system (on/off oxygen), improved the sensitivity and positive predictive value for a similar number of escalations (workload). These findings translate into the following observation level findings (using a threshold of ≥7): The workload was the same (137 alerts per 1000 vital sign observation sets). NEWS-FiO_2 increased the positive predictive value (an additional six adverse event per 1000 alerts) and the sensitivity (an additional 33 alerts per 1000 adverse events). At the admission level (using a threshold of ≥7): there were a total of 83304 admissions in the oxygen therapy cohort. In this cohort NEWS-FiO_2 would have correctly identified an additional 173 individual admissions who went on to have an adverse event.

Comparison to previous studies

To our knowledge, no widely used general adult EWS includes FiO_2 as a predictor variable.(18)(19) Carle et al. designed and internally validated an obstetric EWS using the FiO_2 required to maintain SpO_2 > 96% as a variable.(9) However this EWS has not been translated into widespread use. We adopted the machine learning, decision tree methodology of Badriyah et al., who produced the first decision tree EWS (DTEWS).(20) Ours is the first study to use decision tree analysis for the derivation of thresholds for FiO_2. It is also the first study to use a machine learning method to add a variable to NEWS and evaluate its effect on performance. The relatively modest performance gain achieved by adding FiO_2 is comparable to previous studies that evaluated adding individual vital signs to EWS systems. (21)

Implications for clinicians and policy makers

EWS systems are well established in the UK, with the heuristically developed NEWS being used in 75% of NHS hospitals.(10)(12) Since then, digital EWS platforms have been developed, meaning complex algorithms using vital sign observation sets can be introduced without increasing calculation error. (14)(22) NEWS2 is a new score being adopted nationally in the UK. It is specifically designed to improve EWS performance in patients with hypercapnic

respiratory dysfunction.(10) NEWS2 emphasises the interrelationship between oxygen therapy (or lack thereof) and harm in high risk patient groups. We propose quantifying oxygen therapy via FiO_2 and evaluating the associated relationships with adverse events may be a logical first step in evaluating this important research question.

Limitations

Assumptions in the FiO_2 derivation formula led to some minor but systemic error in the calculation of FiO_2 across the patient cohort. This error is clarified in detail in the supplementary digital content **(SDC-7)**. We also acknowledge that not all patients on high-flow nasal prong oxygen therapy or Non-Invasive Ventilation modes will achieve a FiO_2 of 1.0. However, this will not have affected the score performance because the lower limit for the high scoring FiO_2 band was 53%. This assumption introduced some error to the analysis. Using death or unplanned ICU admission within 24 hours as an outcome measure was in accordance with similar research. However, this outcome measure has limitations. We did not have the data to exclude patients on 'end of life' pathways. Confounding will occur in retrospective, observational data analysis in patient populations where EWS systems are in use. In this study, scoring for oxygen therapy using NEWS increased the risk score, potentially above the alerting threshold. This should have activated a clinical review, potentially facilitating the patient avoiding an adverse event. This trigger in turn becomes a false positive result and reduces the AUROC.(23)(3) Finally, by deriving and testing EWS systems in databases derived from hospitals with EWS in place, the study was seeking to demonstrate incremental gains, which may have been more difficult to detect. A combination of all these factors could explain the modest performance gain seen from NEWS-FiO_2 to NEWS and merit further investigation.

Strengths

Our study is the first to use an automated process such as decision tree analysis to introduce an additional variable to an EWS in a data driven way. We evaluated the effect in a totally separate patient population. We show that a more a granular score for oxygen therapy improves EWS performance. By using the TRIPOD guidelines, we adhered to best practice and ensured the reporting of methods and results are transparent and robust.(24) The recent introduction of NEWS2 makes the timing of this study important.(10)

Future work

Further research is needed to evaluate the relationship between FiO_2 and SpO_2 and their combined associations with adverse events in ward patients, in patients with and without chronic pulmonary disease.

Conclusion

Our study demonstrates that decision tree analysis is an effective method when adding FiO_2 to NEWS in a data driven way. In the ≈ 40% of ward patients requiring oxygen therapy, NEWS-FiO_2 outperformed NEWS when predicting in-hospital death or unplanned ICU admission in the next 24 hours. Adding FiO_2 to NEWS (and other EWS) warrants further study, particularly in patients with or at risk of respiratory dysfunction.

References

1. Smith GB, Prytherch DR, Schmidt PE, et al.: Review and performance evaluation of aggregate weighted “track and trigger” systems. *Resuscitation* 2008; 77:170–179
2. Prytherch DR, Smith GB, Schmidt PE, et al.: ViEWS-Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* 2010; 81:932–937
3. Tarassenko L, Clifton DA, Pinsky MR, et al.: Centile-based early warning scores derived from statistical distributions of vital signs. *Resuscitation* 2011; 82:1013–1018
4. Evans RS, Kuttler KG, Simpson KJ, et al.: Automated detection of physiologic deterioration in hospitalized patients. *J Am Med Inform Assoc* 2015; 22:350–360
5. Beasley R, Aldington S, Robinson G: Is it time to change the approach to oxygen therapy in the breathless patient? *Thorax* 2007; 62:840–841
6. Downs JB: Has oxygen administration delayed appropriate respiratory care? Fallacies regarding oxygen therapy. *Respir Care* 2003; 48:611–20
7. Alam N, Hobbelink EL, van Tienhoven AJ, et al.: The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. *Resuscitation* 2014; 85:587–594
8. Paternina-Caicedo A, Miranda J, Bourjeily G, et al.: Performance of the Obstetric Early Warning Score in critically ill patients for the prediction of maternal death. *Am J Obstet Gynecol* 2017; 216:58.e1-58.e8
9. Carle C, Alexander P, Columb M, et al.: Design and internal validation of an obstetric early warning score: Secondary analysis of the Intensive Care National Audit and Research Centre Case Mix Programme database. *Anaesthesia* 2013; 68:354–367
10. Royal College of Physicians: National Early Warning Score (NEWS) 2 - Standardising the assessment of acute-illness severity in the NHS. Report of a working party. 2017.
11. Royal College of Physicians: National Early Warning Score (NEWS) - Standardising the assessment of acute-illness severity in the NHS. Report of a working party. 2012.
12. Smith GB, Prytherch DR, Meredith P, 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; 84:465–470
13. Collins GS, Altman DG, Moons KGM, et al.: Transparent Reporting of a model for Individual Prognosis Or Diagnosis (to be submitted). *Br Med J* 2014; 68:134–143
14. Wong D, Bonnici T, Knight J, et al.: SEND: A system for electronic notification and documentation of vital sign observations. *BMC Med Inform Decis Mak* 2015; 15:68

15. Bateman NT, Leach RM: Acute oxygen therapy. *Bmj* 1998; 317:798
16. Badriyah T, Briggs JS, Meredith P, et al.: Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS). *Resuscitation* 2014; 85:418–423
17. Munz ED: Psychotherapie in der Psychiatrie. *Nervenheilkunde* 2017; 36:800–805
18. Chapman SM, Wray J, Oulton K, et al.: Systematic review of paediatric track and trigger systems for hospitalised children. *Resuscitation* 2016; 109:87–109
19. Smith MEB, Chiovaro JC, O'Neil M, et al.: Early Warning System Scores for Clinical Deterioration in Hospitalized Patients: A Systematic Review. *Ann Am Thorac Soc* 2014; 11:1454–1465
20. Badriyah T, Briggs JS, Meredith P, et al.: Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS). *Resuscitation* 2014; 85:418–423
21. Watkinson PJ, Pimentel MAF, Clifton DA, et al.: Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data. *Resuscitation* 2018; 129:55–60
22. Prytherch DR, Smith GB, Schmidt P, et al.: Calculating early warning scores-A classroom comparison of pen and paper and hand-held computer methods. *Resuscitation* 2006; 70:173–178
23. Churpek MM, Yuen TC, Edelson DP: Predicting clinical deterioration in the hospital: The impact of outcome selection. *Resuscitation* 2013; 84:564–568
24. Collins GS, Reitsma JB, Altman DG, et al.: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol* 2015; 67:1142–1151

Tables and Figures

Table 1. Demographics and outcomes

Total cohort	OUHNHSFT Training database	QAH Test database
Admissions (total)	42764	217363
Admissions (with > 0 complete vital sign observation set)	29931 (69.9%)	120017 (55.2%)
Admissions with an event outcome (%)	1669 (3.9%)	7523 (3.5%)
Admissions (male) (%)	14887 (49.7%)	56140 (46.7%)
Admission age, mean (SD)	64(19)	63(20)
Vital sign observation sets	1201714	5545039
Vital sign observation sets tagged as event outcome (%)	9412(0.8%)	42653 (0.8%)
Vital sign observation sets tagged as unplanned ICU outcome (%)	5503 (0.5%)	15029 (0.3%)
Vital sign observations sets tagged as death outcome (%)	3909 (0.3%)	27621 (0.5%)
Length of stay, median (IQR)(hours)	100 (170)	83 (143)
Heart rate, mean (SD) (beats per minute)	82 (16)	80 (16)
Respiratory rate, mean (breaths per minute)	17 (3)	17 (3)
Systolic blood pressure, mean (SD) (mmHg)	127 (22)	126 (22)
FiO_2, mean (SD)(%)	26 (15)	26 (12)
Body temperature, mean (degrees Centigrade)	36.4 (0.6)	36.7 (0.5)
Oxygen cohort		
Admissions (% of total cohort)	17012 (39.7%)	83304 (38.3%)
Admissions with an event outcome (% of oxygen cohort)	1027 (6.0%)	5688 (6.8%)
Admissions (male) (% of oxygen cohort)	8166 (48%)	37487 (45%)
Admission age, mean (SD) (years)	65 (18)	68 (17)

Vital sign observation sets	222156 (18.4%)	1055189 (19%)
Vital sign observation sets tagged as event outcome (%)	6469 (2.9%)	30359 (2.8%)
Vital sign observation sets tagged as unplanned ICU outcome (%)	2443(1.1%)	9843 (0.9%)
Vital sign observations sets tagged as death outcome (%)	4026(1.8%)	20516 (1.9%)
Length of stay, median (IQR)(hours)	131 (220)	131 (213)
Heart rate, mean (SD) (beats per minute)	85 (18)	84 (18)
Respiratory rate, mean (SD)(breaths per minute)	18 (4)	18 (4)
Systolic blood pressure, mean (SD) (mmHg)	125 (23)	125 (23)
FiO_2, mean (SD)(%)	49 (18)	47 (15)
Body temperature, mean (SD)(degrees Centigrade)	36.4 (0.7)	36.7 (0.6)

Table 1. *Demographic descriptors for the admissions included in each of the OUHNHSFT Training and QAH Test databases.*

Table 2. FiO_2 scoring bands statistics in the QAH Test database for patients receiving oxygen therapy

Score	0	1	2	3	Sum
FiO_2 thresholds (%)	21 – 22	22.1 - 37	37.1 - 53	> 53	
Vital sign observation sets	3883	234504	564712	252090	1055189
Vital sign observation sets tagged as an event	56	6779	9066	14458	30359

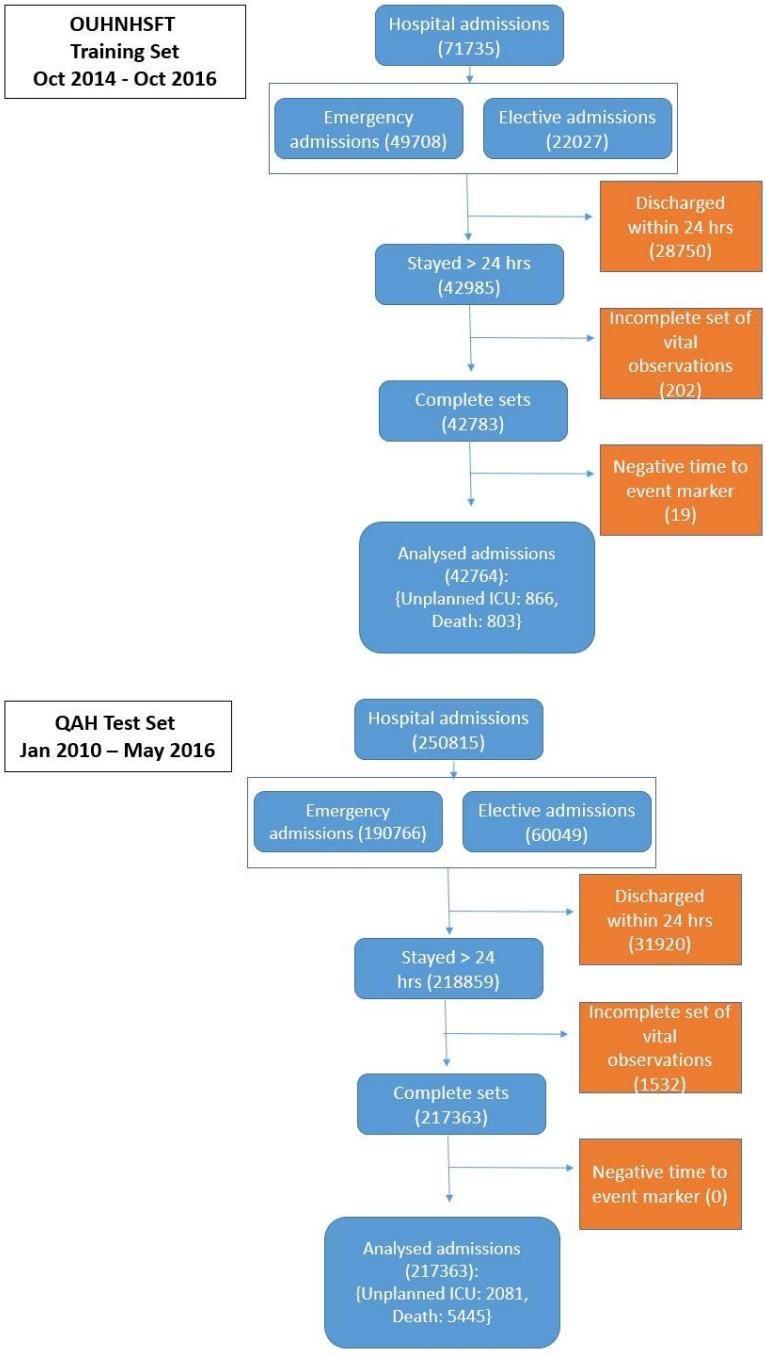
Table 2. *Analysis on the FiO_2 scoring bands statistics. The table summarises how oxygenated patient observations are clustered in each of the proposed FiO_2 bands.*

Table 3. Performance of NEWS and NEWS-FiO_2 in oxygen requiring cohort in the QAH Test database

	NEWS	NEWS-FiO_2
AUROC (CI)	0.811 (0.809 - 0.814)	0.823 (0.819 - 0.824)
Sensitivity (%) (Score ≥5 / ≥7)	81.4/56.9	82.7/60.2
Specificity (%) (Score ≥5 / ≥7)	64.7/87.5	64.8/87.6
Positive Predictive Value (%) (Score ≥5 / ≥7)	6.4/11.9	6.5/12.5
Efficiency (%) (Score ≥5 / ≥7)	36.6/13.7	36.5/13.7

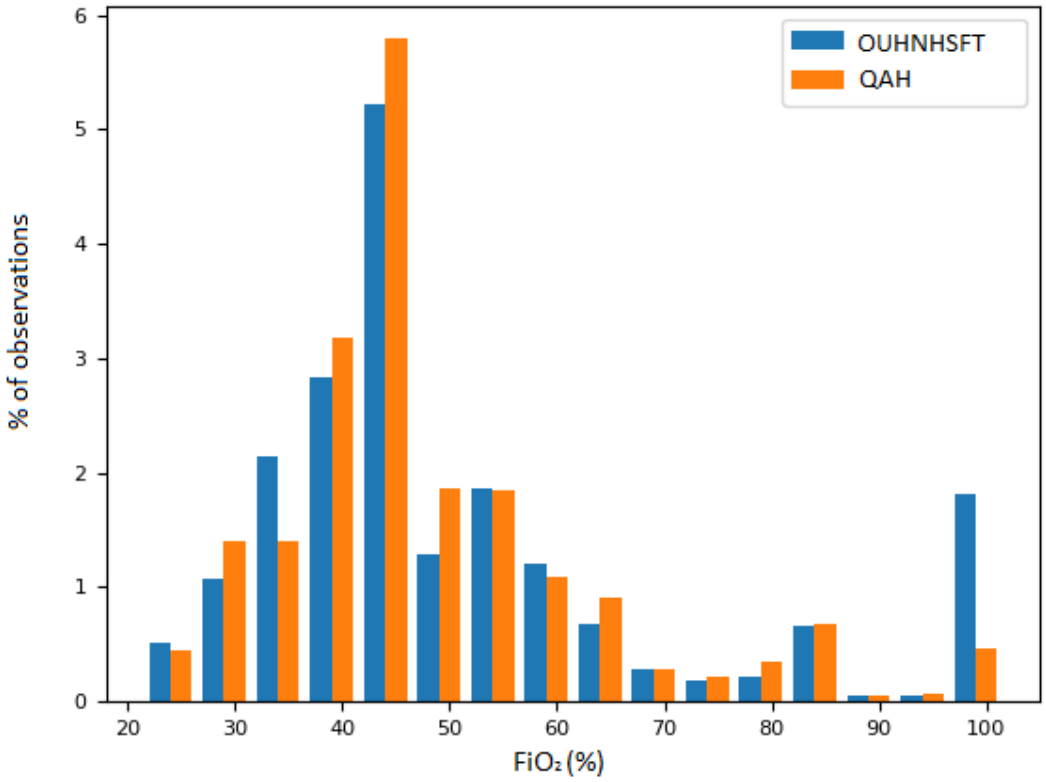
Table 3*. Performance metrics of the scoring systems (NEWS, NEWS-FiO_2) for predicting the event outcome in the QAH Test database, which includes the Area Under the Receiver Operating Characteristic curve (AUROC), with 95% confidence interval (CI), and sensitivity, specificity and positive predictive value values at a threshold of 5 and 7.*

Figure 1.



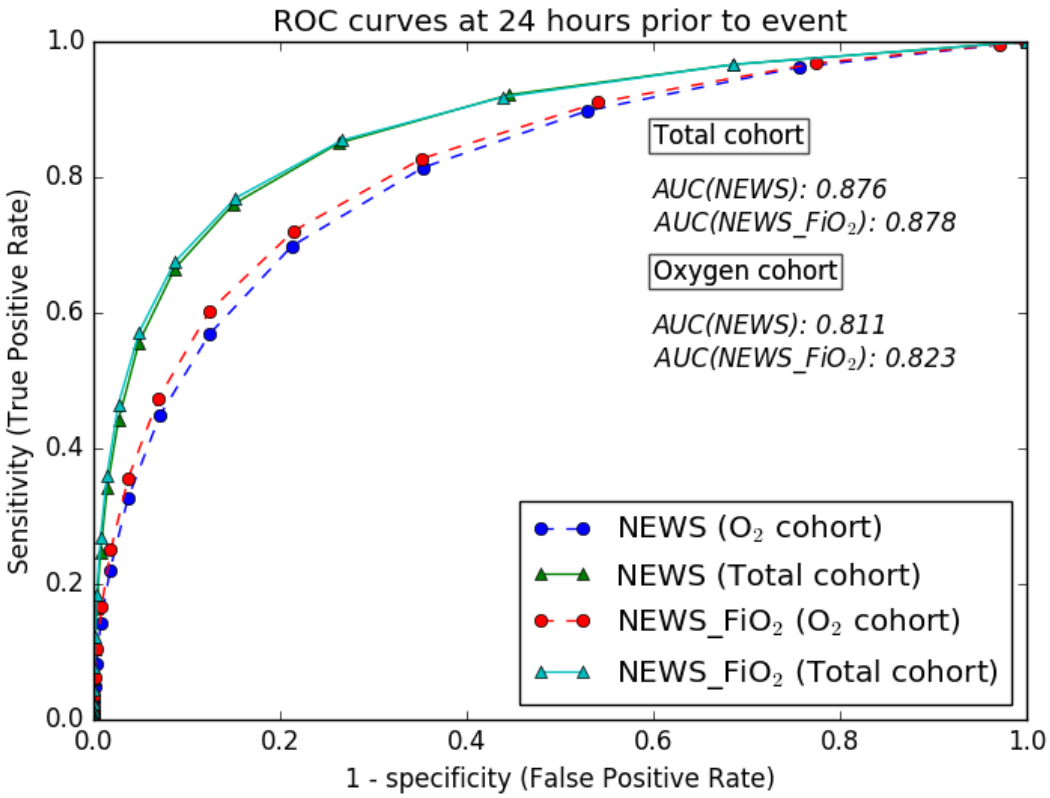
(1) Flowchart of included and excluded patients in both databases

Figure 2.



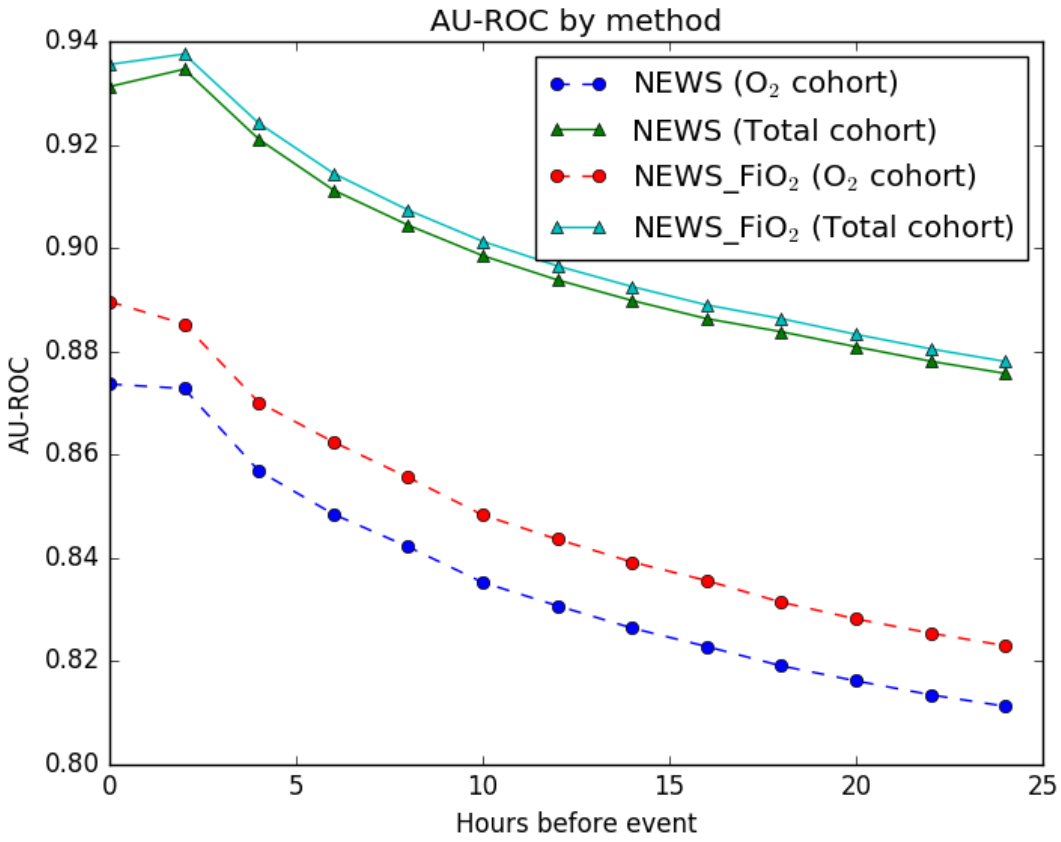
(2) FiO_2 histogram of the oxygen requiring cohort in both databases. QAH: Queen Alexander Hospital, OUHNHSFT: Oxford University Hospitals National Health Service Foundation Trust. Percentage of total observations (both oxygen requiring and non-oxygen requiring vital-sign observations sets), divided into %% bands from 35 – 100%

Figure 3a.



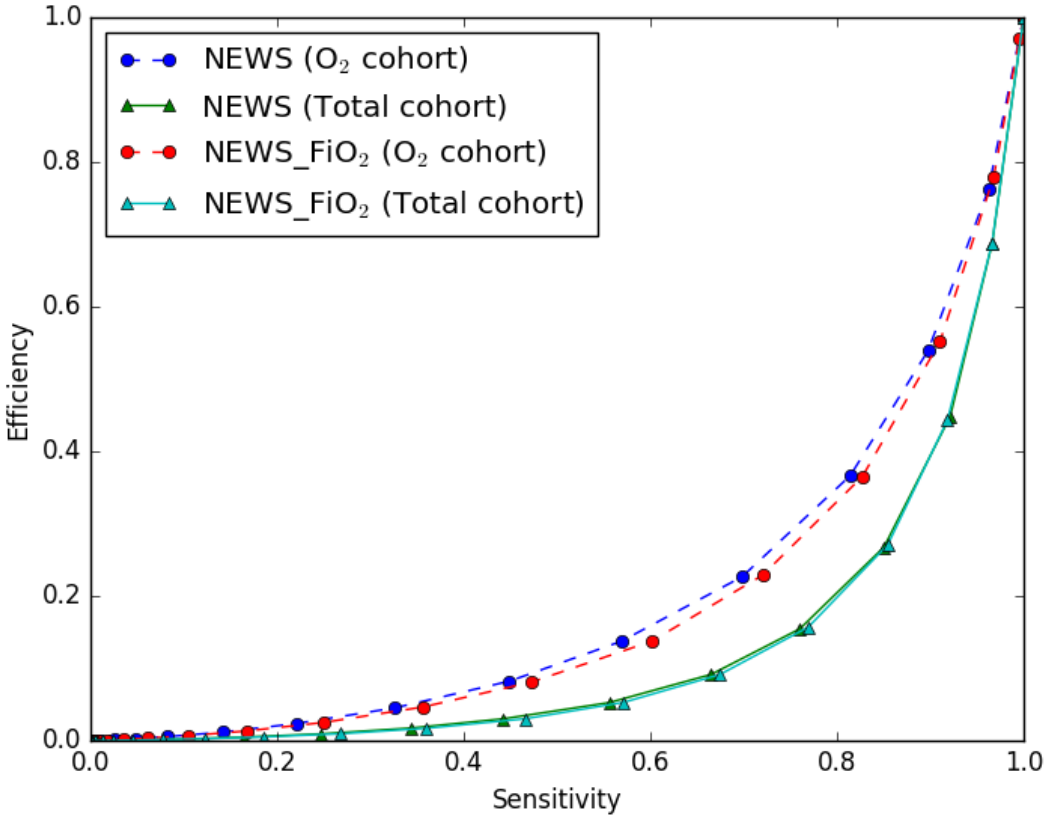
(3a) Receiver operating characteristic curve (ROC) for NEW and NEWS-FiO_2 in oxygen requiring and total patient cohorts.

Figure 3b.



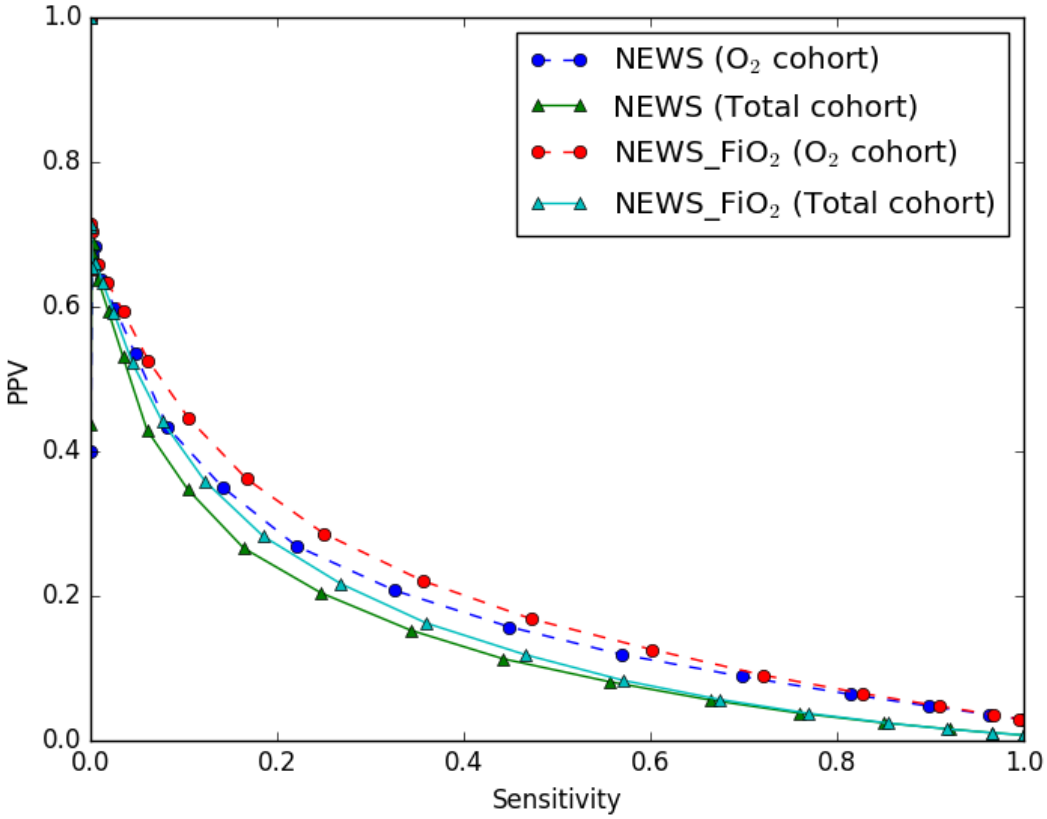
(3b) Area Under the Operating Characteristic curve (AU-ROC) performance when time-to-event approaches the event time for NEW and NEWS-FiO_2 in oxygen requiring and total patient cohorts

Figure 3c.



(3c) Efficiency curves for NEWS and NEWS-FiO_2 in oxygen requiring and total patient cohorts. The curve shows the fraction of the total number of observations at, or above, each EWS value against the fraction of the total number of observations for which the event outcome was true at, or above that EWS value

Figure 3d.



(3d) Precision-recall curves for NEWS and NEWS-FiO_2 in the oxygen requiring and total patient cohorts.

Supplementary Material

SDC-1 TRIPOD checklist

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D; V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	title
Abstract	2	D; V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	abstract
Introduction				
Background and objectives	3a	D; V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	Intro .P2
	3b	D; V	Specify the objectives, including whether the study describes the development or validation of the model or both.	Intro .P3
Methods				
Source of data	4a	D; V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	Meth.P1
	4b	D; V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	Src Data .P1
Participants	5a	D; V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centers.	Part. P1

	5b	D; V	Describe eligibility criteria for participants.	Part. P2
	5c	D; V	Give details of treatments received, if relevant.	NA
Outcome	6a	D; V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	Outcomes
	6b	D; V	Report any actions to blind assessment of the outcome to be predicted.	NA
Predictors	7a	D; V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	Pred
	7b	D; V	Report any actions to blind assessment of predictors for the outcome and other predictors.	FiO 2.P2
Sample size	8	D; V	Explain how the study size was arrived at.	SDC -3
Missing data	9	D; V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	Missing Data
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	DT
	10 b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	DT
	10c	V	For validation, describe how the predictions were calculated.	DT & FiO 2
	10 d	D; V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	Eval .P1

	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	NA
Risk groups	11	D; V	Provide details on how risk groups were created, if done.	Eval .P2
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	NA
Results				
Participants	13a	D; V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Res. P1
	13 b	D; V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Res. P1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	Res. P2
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	Res. P1,2
	14 b	D	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Res. P3
	15 b	D	Explain how to the use the prediction model.	Res. P3

Model performance	16	D; V	Report performance measures (with CIs) for the prediction model.	Perf. P1
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	NA
Discussion				
Limitations	18	D; V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	*Limit*
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	Dis. P1,2,3 + Comp
	19b	D; V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	Dis. P1,2,3 + Stren
Implications	20	D; V	Discuss the potential clinical use of the model and implications for future research.	DisP,Impl
Other information				
Supplementary information	21	D; V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	*Suppl*
Funding	22	D; V	Give the source of funding and the role of the funders for the present study.	*Fund*

SDC-2 GCS (Glasgow Coma Scale) converted to AVPU (Alert Verbal Pain Unresponsive) scale following the table below.

GCS	**AVPU**
>=15	1 (Alert)
12 - 15	2 (Verbal)
8-12	3 (Pain)
<=8	4 (Unresponsive)

GCS to AVPU conversion parameters.

SDC-3 Mask types and corresponding flow rates

For the documentation of the oxygen delivery system, the device codes within both NHS trusts are aligned according to BTS guideline recommendations. (23)

BTS codes	**OUHT database**		**Mask Prevalence (100)**	**PHT database**		**Mask Prevalence (%)**
	mask name	**coded value (%)**		**mask name**	**Coded value (%)**	
A	Room Air	21		Room Air	21	
V	Venturi oxygen 24%	24	1.6	Venturi (variable)	R/Rand***	11.2
V	Venturi oxygen 28%	28	1.8			
V	Venturi oxygen 35%	35	1.3			
V	Venturi oxygen 40%	40	0.6			
V	Venturi oxygen 60%	60	0.3			
H	Humidified oxygen 28%	28	0.7			
H	Humidified oxygen 35%	35	0.5			
H	Humidified oxygen 40%	40	0.3			

H	Humidified oxygen 60%	60	0.4	Humidified (60% max)	60	4.3
H	Humidified oxygen 80%	80	0.1			
H	Humidified oxygen 98%	98	0.1	Humidified (98% max)	98	0.0
N	Nasal cannula	converted*	71.4	Nasal cannula	converted*	72.6
SM	Simple mask	converted*	11.6	Simple Mask	converted*	7.0
RM	Reservoir mask	80	2.2	Reservoir	80	2.7
TM	Tracheostomy mask	converted*	0.3	Tracheostomy / Trach flow	R/converted*	0.4
CP	$CPAP^{+}$	100	0.6	$CPAP^{+}$ / $BIPAP^{++}$	100	1.1
NIV	Non-invasive	100	1.6	NIV^{+++}	100	0.4
OTH	Other device	converted*	0.0	Aerosol	converted*	0.3
SM	Nebuliser Mask	converted*	0.03	Nasal humidified	converted*	0.0
HFN	High Flow / Optiflow	100	4.7	Intubated Conc./Flow	100	0.002

Mask types and corresponding flow rates in percentage, from each trust: Oxford University Hospitals Trust (OUHT) and Portsmouth Hospitals Trust (PHT).

** converted refers to the use of the Bateman equation for computing the FiO_2 value.*

*** R/converted refers to conditions where the O_2 flow rate has been recorded in a mixed and potentially inaccurate manner. In these circumstances, we made the assumption that O2 flows recorded > 21 were in fact FiO2 recordings and we analysed those values as percentage FiO2*

values. For any values recorded as < 21, we used the Bateman equation for converting the O_2 Flow rate to FiO_2.

*R/Rand*** refers to conditions where the O_2 flow rate has been recorded in a mixed and potentially inaccurate manner. In these circumstances, we first checked whether the value reported as O_2 flow rate was > 21. If so, the value was considered as the mistakenly reported FiO_2 and analysed as such. For any values < 21, we randomly generated a percentage in the valid range for Venturi masks (24%-60%).*

[+] *CPAP: Bilevel Positive Airway Pressure mask*

[++] *BIPAP: Continuous Positive Airway Pressure mask*

[+++] *NIV: Non-Invasive Ventilation mask*

SDC-4: Error rates for different flow rates and tidal volumes

Ratio of O_2 Flow rate to Respiratory Rate (R: liter/breath)	**Tidal Volume (liter)**	**FiO_2 (%)**
	0.4	1.00
0.5	0.45	1.00
	0.6	0.87
	0.4	1.00
0.4	0.45	0.91
	0.6	0.74
	0.4	0.80
0.3	0.45	0.74
	0.6	0.60
	0.4	0.60
0.2	0.45	0.56
	0.6	0.47
	0.4	0.41
0.1	0.45	0.39
	0.6	0.34
	0.4	0.23
0.01	0.45	0.23
	0.6	0.22

*Error rates in calculating FiO_2. O_2 flow rate/respiratory rate ratio values are combined with different tidal volume assumptions (0.45, 0.4 and 0.6) to produce different Fi02 values - FiO2 = (0.21 +0.79*R/TV). We used a fixed tidal volume of 0.45 (liter per breath) for all patients. The results show the variation in FiO_2 for patients weighing between 57kg (Vt of 0.4) and 85kg (Vt of 0.6) based and a tidal volume of 7mls/kg. The variation in FiO_2 will increase the score by one unit in R values equal to 0.2 and 0.1 extreme case of TV=0.6.*

SDC-5 Comparison of scoring systems

Vitals	**EWS**	**3**	**2**	**1**	**0**	**1**	**2**	**3**
RR	**NEWS**	**< 8**		**9 - 11**	**12 - 20**		**21 - 24**	**> 24**
HR	**NEWS**	**<40**		**41-50**	**51-90**	**91-110**	**111-130**	**>131**
SBP	**NEWS**	**<90**	**91 -100**	**101-110**	**111-219**			**>219**
SpO_2	**NEWS**	**<91**	**92-93**	**94-95**	**>96**			
Temp	**NEWS**	**< 35**		**35.1 - 36**	**36.1 - 38**	**38.1 - 39**	**>39.1**	
FiO_2	**NEWS**		**Yes**		**No**			
	NEWS-FiO_2				**0-22**	**22.1-37**	**37.1-53**	**>53**
AVPU	**NEWS**				**A**			**V,P,U**

NEWS: National Early Warning Score Decision Tree Early Warning Score of Oxford, where the scoring bands are derived using the OUHNHSFT Training dataset collected at the Oxford University Hospitals Trust.

SDC-6 Decision Tree Algorithm

The decision tree is a graphical model that is composed of a set of nodes and edges which are organised in hierarchical structure. The tree originates from a root node which contains the whole data. The distribution of the labelled samples (observations recorded within the 24 hours preceding an event) at the root node are divided into two subsets, using a decision function. The aim is to search for the best feature and the best split in that feature's range which maximizes the data purity in the subsequent subsets. Denoting the training data as $(\mathbf{x}_i; y_i) = (x_1, x_2, \ldots, x_d, y_i)$, a decision tree recursively partitions the space such that the samples with the same labels are grouped together. Representing the data at node n by Q. For each candidate $split = (j, t_m)$ consisting of a feature j and threshold t_m, the algorithm partitions the data into $Q_{left}(\theta): (\mathbf{x}; y)|x_j <= t_m$ *and* $Q_{right}(\theta) : Q \backslash Q_{left}(\theta)$ subsets. The impurity at node n can then be formulated as:

$G(Q, \theta) = (n_{left}/N_n) H(Q_{left}(\theta)) + (n_{right}/N_n) H(Q_{right}(\theta))$

where $H(Xn) = \Sigma k\, p_n k\, (1 - p_n k)$ is known as the Gini impurity measure and k denotes the number of classes. The aim of the partitioning is then to select the parameters that minimises the impurity:

$\theta^* = argmin\ \theta\ G(Q, \theta)$

The algorithm recursively partitions data in hierarchical fashion for subsets $Q_{left}(\theta^*)$ and the algorithm recursively partitions data in hierarchical fashion for subsets $Q_{left}(\theta^*)$ and $Q_{right}(\theta^*)$ until the maximum allowable depth or the minimum number of node samples is reached (this is set to ten samples in our implementation). This in practice in a uni-variate tree, infers that the same feature will be recursively partitioned at different thresholds until one of the stopping criteria is reached. Once a decision tree is formed for a vital sign data i.e. RR, each node in the tree is assigned with a risk value that is estimated as the proportion of the number of observations followed by an abnormal event to the total number of observations reaching that node. The conversion of the node risk values to weighting scores is then undertaken following Algorithm1: where the nodes with a node risk value less than mean risk (μr), was ascribed the value 0 , the node risk values greater than the mean risk and less than two times the mean risk were ascribed value 1, the node risk values greater than two times the mean risk and less than three times the mean risk are ascribed value 2 and finally if the node risk value is greater than three times the mean risk, it was ascribed to value 3.

Algorithm 1: DTEWS weight association algorithms

Require: μ_r -> $|N^{event}|\backslash|N^{Total}|$

Require: treeRisks -> {nodeRisk(n)}, n N ode where N ode is set of tree nodes and nodeRisk(n) = $|N_n^{event}|\backslash|N_n^{Total}|$ with

N_n representing the number of samples at node n in the tree.

for i, nodeRisk in enumerate(treeRisks):

if (nodeRisk < μ_r):

nodeScore[i] = 0

else if (μ_r < nodeRisk < 2 μ_r):

nodeScore[i] = 1

else if (2 μ_r <nodeRisk < 3 μ_r):

nodeScore[i] = 2

else if (nodeRisk> 3 μ_r):

nodeScore[i] = 3

end

SDC-7: FiO_2 scoring algorithm

Step 1. Align the trust-dependent Oxygen mask coding system with the BTS standard

Step 2. Use the coding system introduced in SDC-5 to compute a mask-independent FiO_2 value in percentage

Step 3. Use the FiO_2 thresholding values (shaded in green) introduced in SDC-3 Table to compute an EWS score for each recorded value

Step 4. Aggregate the score computed for the FiO_2 observations of Step 3 with scores computed for all other vital observations in each EWS-FiO_2 model (NEWS-FiO_2 and DTEWSO-FiO_2)

Chapter 3. Evaluating HAVEN in the ward environment

Multiple prognostic algorithms have been developed using retrospective, data driven validation techniques, however few have undergone 'real world' ward-based evaluation. This retrospective validation strategy is supported by the current international Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines but it does have limitations. EWS systems perform differently prospectively when compared to retrospective validation study results (often worse). This is because of human error and because other unpredictable elements in the ward environment affect the way clinical staff care for patients. To provide more insight into this issue, an embedded mixed-methods study was conducted. The primary objective of the study was to identify factors worsening HAVEN performance. Participants were identified using a digital user interface that displayed real-time HAVEN risk scores for all currently admitted John Radcliffe Hospital patients. Methods of constant comparison (i.e., thematic analysis) were used to analyse the structured patient data extracted from the EPR regarding patients with high HAVEN scores. The results of the study showed the protype HAVEN algorithm generated a high number of false positive alerts. Qualitative analysis showed this misclassification occurred in three main groups: firstly, patients with objective markers of clinical instability but who responded to ward-based measures, secondly patients with objective markers of clinical instability but who were referred to other (non-ICU) specialist teams and, thirdly, patients who were elderly and frail with Treatment Limitations in place. These findings were used to iterate the HAVEN algorithm and improve performance during development.

TESTING A DIGITAL SYSTEM THAT RANKS THE RISK OF UNPLANNED INTENSIVE CARE UNIT ADMISSION IN ALL WARD PATIENTS: PROTOCOL FOR A PROSPECTIVE OBSERVATIONAL COHORT STUDY

Authors

James Malycha MBBD FCICM[1,2,3], Oliver Redfern MBBS PhD[1], Guy Ludbrook PhD MBBS FANZCA[2], Duncan Young BM DM FRCA FMedSci[1], Peter Watkinson MD ChB MRCP FDICM[1]

Affiliations

[1]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[2]Discipline of Acute Care Medicine, University of Adelaide, South Australia, Australia

[3]Intensive Care Unit, The Queen Elizabeth Hospital, Adelaide, South Australia, Australia

Keywords

Clinical deterioration, intensive care unit, critical care unit, predictive score, electronic patient record, qualitative medical note review

Funding

This publication presents independent research funded by the Health Innovation Challenge Fund (HICF-R9-524; WT-103703/Z/14/Z), a parallel funding partnership between the Department of Health and Wellcome Trust.

Ethics

Approval has been granted for gathering the data used in this study (South Central – Oxford C, Research Ethics Committee) REC reference: 16/SC/0264, 13th June 2016) and Confidentiality Advisory Group (16/CAG/0066).

Statement of Authorship

Title of paper	Testing a digital system that ranks the risk of unplanned Intensive Care Unit admission in all ward patients: protocol for a prospective observational cohort study
Publication status	Published
Publication details	Malycha J, Redfern OC, Ludbrook G, et al.: Testing a digital system that ranks the risk of unplanned intensive care unit admission in all ward patients: Protocol for a prospective observational cohort study. *BMJ Open* 2019; 9:1–4

Principal Author

Name of Principal Author	Dr James Malycha
Contribution to paper	Conceptualised the work, wrote manuscript and was corresponding author
Overall percentage (%)	90%
Certification	This paper reports on original research conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Co-author contributions

Each author certifies:

- The candidate's stated contribution to the publication is accurate (as state above);
- Permission is granted for the candidate to include the publication in the thesis; and
- The sum of all co-author contributions is equal less the candidate's stated contribution

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Oliver Redfern
Contribution to paper	Assisted with design of the study, methodological planning and evaluated and edited the manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor David Clifton
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Guy Ludbrook
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Duncan Young
Contribution to paper	Assisted with study design, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Peter Watkinson
Contribution to paper	Assisted with study design, evaluated and edited manuscript

Abstract

Introduction

Traditional early warning scores (EWS) use vital sign derangements to detect clinical deterioration in patients treated on hospital wards. Combining vital signs with demographics and laboratory results improves EWS performance. We have developed the Hospital Alerting Via Electronic Noticeboard (HAVEN) system. HAVEN uses vital signs, as well as demographic, comorbidity and laboratory data from the electronic patient record (EPR), to quantify and rank the risk of unplanned admission to an Intensive Care Unit (ICU) within 24 hours for all ward patients.

The primary aim of this study is to find additional variables, potentially missed during development, which may improve HAVEN performance. These variables will be sought in the medical record of patients misclassified by the HAVEN risk score during testing.

Methods

This will be a prospective, observational, cohort study conducted at the John Radcliffe Hospital, part of the Oxford University Hospitals National Health Service Foundation Trust in the United Kingdom.

Each day during the study periods, we will document all highly ranked patients (i.e. those with the highest risk for unplanned ICU admission) identified by the HAVEN system. After 48 hours we will review the progress of the identified patients. Patients who were subsequently *admitted* to the ICU will be removed from the study (as they will have been correctly classified by HAVEN). Highly ranked patients *not admitted* to ICU will undergo a structured medical notes review. Additionally, at the end of the study periods, all patients who had an unplanned ICU admission but whom HAVEN *failed to rank highly* will have a structured medical notes review. The review will identify candidate variables, likely associated with unplanned ICU admission, not included in the HAVEN risk score

Discussion

Our study will use a clinical expert conducting a structured medical notes review to identify variables, associated with unplanned ICU admission, not included in the development of the

HAVEN risk score. These variables will then be added to the risk score and evaluated for potential performance gain. To our knowledge, this is the first study of this type. We anticipate that documenting the HAVEN development methods will assist other research groups developing similar technology.

Strengths and limitations of the study

- The study methodology is in accordance with the STROBE guidelines
- We describe a method that combines risk score testing with a structured medical notes review conducted by a clinical expert for the iterative improvement of a digital system that quantifies risk for unplanned Intensive Care Unit admission in all ward patients
- To our knowledge, this is the first study of this type

Background

Introduction

Early Warning Score (EWS) systems, such as the National Early Warning Score (NEWS), combine abnormalities in patient vital signs into an aggregate score. (1) This score triggers a clinical response when a threshold is exceeded. Despite wide-scale adoption of EWS systems, significant clinical patient deterioration on hospital wards still occurs. (1, 2) Additionally, high numbers of false alerts lead to alert 'fatigue' and inefficient use of response teams. (3) Adding additional clinical information to such systems, such as laboratory results and co-morbidities, improves specificity. (4–12) However, identifying and adding new variables requires a systematic approach to avoid needless complexity. (13)

We have developed a system to predict the risk of unplanned ICU admission (within 24 hours) for patients on general medical and surgical wards. It is called Hospital Alerting Via Electronic Noticeboard (HAVEN). (14) To identify potential variables for inclusion in HAVEN, we used a modified Delphi process and a systematic literature review. (15) Those identified variables that were available within the Electronic Patient Record (EPR) were extracted from datasets comprising all patients admitted to two National Health Service (NHS) Trusts (a Trust is a legal entity that provides goods and services for the purposes of the provision of hospital, community and/or other aspects of patient care). (12) We then used a machine learning method (16) to select the optimal combination of variables for the HAVEN risk score. In contrast to EWS systems, HAVEN was not designed to produce alerts. Instead, HAVEN provides a list of patients in the hospital, ranked from most to least at risk of requiring an ICU admission. The intent is that HAVEN will improve patient safety by informing the use of clinical response teams.

Aims and Objectives

The primary aim of this study is to discover additional candidate variables, not recognised during the data driven derivation process that would improve the performance of the HAVEN risk score. We will review the medical records of *misclassified* patients, i.e. patients ranked highly by HAVEN but who were not admitted to the ICU; or patients who were never ranked highly by HAVEN but had an unplanned ICU admission.

The HAVEN risk score

The HAVEN risk score is calculated using both *static* and *dynamic* variables extracted in real-time from the EPR.

Static variables refer to patient-level data available at admission: age, gender, co-morbidities (classified according to the Elixhauser Co-morbidity index (17)), and Hospital Frailty Risk Score. (18) As diagnostic coding in the United Kingdom (UK) occurs after a patient has been discharged, the co-morbidity index and frailty scores are calculated using a patient's admissions over the previous two years. Score performance in patients with no previous admissions (and potentially undocumented comorbidities) will be evaluated separately.

Dynamic variables refer to measurements taken repeatedly during hospital admission, i.e. laboratory results and vital signs. The HAVEN risk score is currently updated according to the most recent measurements of: albumin, bilirubin, C-reactive protein (CRP), haemoglobin, platelets, white cell count, potassium, sodium, urea, creatinine, heart rate, systolic blood pressure, respiratory rate, body temperature, a neurological status assessment using either the Alert-Verbal-Painful-Unresponsive (AVPU) scale or the Glasgow Coma Scale (GCS), peripheral oxygen saturation from pulse oximetry (SpO_2) and the estimated fraction of inspired oxygen. (19) A patient's HAVEN score is re-calculated each time a new dynamic variable is received by the system and the score is further adjusted for the time since hospital admission.

Methods

The study will be reported according to the Strengthening Reporting of Observational Studies in Epidemiology (STROBE) guidelines. (20)

Design and setting

This is a prospective, observational, cohort study conducted in the John Radcliffe Hospital, part of Oxford University Hospitals National Health Service Foundation Trust in the UK. The John Radcliffe Hospital is a tertiary hospital with over 800 beds and serves a population over 650,000 people, who are generally more affluent and with higher life expectancy than the national average. (21)

Data Collection

Data collection will occur during four, full, non-consecutive weeks in 2019. The notes review will be undertaken by a senior critical care physician. Patients who are discharged or die during the study period will have these details recorded. They will remain in the analysis dataset.

Participants

Eligibility criteria

Emergency and elective adult patients (16 years or over) admitted to medical, surgical, observational or short stay wards will be eligible for inclusion. We will exclude patients for whom a score cannot be generated (i.e. those with no recorded vital sign or laboratory measurements).

Sample size

We will sample two sub-groups of patients:

1. False High Rank (FHR)
2. False Low Rank (FLR)

The False High Rank (FHR) group will consist of patients ranked highly by HAVEN but who were not admitted to the ICU. To identify this group, we will record the five highest ranked patients on the HAVEN system at 9am each morning of the study. After 48 hours, we will remove any patients who were subsequently admitted to the ICU. The remaining patients' records will be reviewed.

The False Low Rank (FLR) group will be identified at the end of the study and consist of all patients who had an unplanned ICU admission during the study period and were not present in any of the daily high-ranking groups. These patients' records will also undergo a medical notes review.

The study will run for four non-consecutive weeks with an expected recruitment of between 130 and 150 patients.

Structured medical notes review

We will carry out a structured review of patient medical notes (electronic and paper-based) for the two sample groups described above. From these, we will construct a medical summary, looking specifically at patient-centred and system-based variables associated with decisions around ICU admission. We will use a modified version of the Hogan el al. qualitative note review techniques. (22) We will then conduct a thematic analysis of the extracted data. (23) It is expected that from within the themes the additional variables will be identified. Along with the as yet unknown variables, the following data will be extracted:

1. Primary diagnosis
2. Comorbidities and past medical history (where not available from previous admissions)
3. Any treatment limitations put in place and the reasons for these including "Do not attempt resuscitation" (DNAR) documents
4. Current medication
5. Radiological imaging
6. Point-of-care blood gas analysis
7. Clinical Frailty Score. (24)

Qualitative methods

Qualitative data (e.g., information in free text) will be analysed thematically, using methods of constant comparison. (25) A coding framework will be constructed to assist understanding of the data. We will use Nvivo Software (QSR International Pty Ltd, **www.qsrinternational.com**) to support the qualitative analysis process.

Patient safety and public involvement

As an observational study of patient records with no intervention, adverse events related to research interventions are not possible. In the event that inadequate care is identified during the structured medical note review, local NHS Trust protocols will be followed. Reviewers will act in accordance with the General Medical Councils Good Medical Practice Guidelines (2013). This action includes acting immediately if a patient is not receiving basic care to meet their needs. If patients are at risk because of inadequate premises, equipment or other resources, policies or systems, we will correct the matter if possible and raise our concerns in line with

workplace policy. All measures will be documented as per local policies. The HAVEN project has had two lay members on the management committee throughout. They have been involved in regular discussions regarding the aims and remit of the HAVEN project.

Discussion

Main findings

This study will use structured medical notes review on ward patients misclassified by HAVEN to identify variables that may enhance performance. Any identified variables will be systematically introduced into our score development pipeline to evaluate whether they improve score performance.

Strengths and limitations of the study

This study is part of a project-wide process to document the development of the HAVEN system such that it is thorough, transparent, repeatable, reportable and the methodology could be useful for other groups developing similar technology.

Unplanned ICU admission is an outcome measure subject to bias, such as the decision-making of individual physicians, local practice guidelines and bed availability. (26, 27) This study is limited to one hospital and the results may not be generalisable to other hospitals. Variables identified from the thematic analysis may not be available in the EPR and therefore cannot risk score be used to improve the performance of the HAVEN risk score. Likewise, patients with no previous admissions to the John Radcliffe Hospital will have no available comorbidity data, potentially limiting performance of the risk score in these patients. To assess the impact of these missing data, we will undertake sub-groups analyses in those patients with/without prior admissions.

While a significant proportion of ICU admissions are referred directly from the Emergency Department (ED), the HAVEN system was designed specifically for ward patients needing the attention of the critical care team. By excluding these ED referrals we are reducing the number of eligible patients for this study.

Implications

To our knowledge, this is the first protocol to describe a study of this type. We hope this protocol it will assist future development of similar systems.

References

1. Royal College of Physicians (London): National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. 2017.
2. Care Quality Commission: Opening the door to change NHS safety culture and the need for transformation. 2018.
3. Curry J, Jungquist CR: A critical assessment of monitoring practices, patient deterioration, and alarm fatigue on inpatient wards: a review. *Patient Saf Surg* 2014; 8:29
4. Churpek MM, Yuen TC, Park SY, et al.: Using Electronic Health Record Data to Develop and Validate a Prediction Model for Adverse Outcomes in the Wards*. *Crit Care Med* 2014; 42:841–848
5. Kang MA, Churpek MM, Zadravecz FJ, et al.: Real-time risk prediction on the wards: A feasibility study. *Crit Care Med* 2016; 44:1468–1473
6. Alvarez CA, Clark CA, Zhang S, et al.: Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. *BMC Med Inform Decis Mak* 2013; 13:28
7. Bailey TCC, Chen Y, Mao Y, et al.: A trial of a real-time Alert for clinical deterioration in Patients hospitalized on general medical wards. *J Hosp Med* 2013; 8:236–242
8. Escobar GJ, Laguardia JC, Turk BJ, et al.: Early detection of impending physiologic deterioration among patients who are not in intensive care: Development of predictive models using data from an automated electronic medical record. *J Hosp Med* 2012; 7:388–395
9. Hackmann G, Chen M, Chipara O, et al.: Toward a two-tier clinical warning system for hospitalized patients. *AMIA . Annu Symp proceedings AMIA Symp* 2011; 2011:511–9
10. Tam V, Frost SA, Hillman KM, et al.: Using administrative data to develop a nomogram for individualising risk of unplanned admission to intensive care. *Resuscitation* 2008; 79:241–248
11. Kipnis P, Turk BJ, Wulf DA, 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; 64:10–19
12. Redfern OC, Pimentel MAF, Prytherch D, et al.: Predicting in-hospital mortality and

unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: Development and validation of a multivariable model. *Resuscitation* 2018; 133:75–81

13. Collins GS, Reitsma JB, Altman DG, et al.: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol* 2015; 67:1142–1151
14. Watkinson, P., Young, J. D., Prytherch, D., Tarassenko, L., Clifton, D., & Briggs J: Hospital Alerting Via Electronic Noticeboard (HAVEN) Study Protocol [Internet]. 2016; Available from: https://ora.ox.ac.uk/objects/uuid:322561c7-866e-4c
15. Malycha J, Bonnici T, Clifton DA, et al.: Patient centred variables with univariate associations with unplanned ICU admission: a systematic review. *BMC Med Inform Decis Mak* 2019; 19:98
16. Chen T, He T, Benesty M, et al.: Extreme Gradient Boosting Package 'xgboost.' 2018; 54
17. Quan H, Sundararajan V, Halfon P, et al.: Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005; 43:1130–9
18. Gilbert T, Neuburger J, Kraindler J, et al.: Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet* 2018; 391:1775–1782
19. Malycha J, Farajidavar N, Pimentel MAF, et al.: The effect of fractional inspired oxygen concentration on early warning score performance: A database analysis. *Resuscitation* 2019; 139:192–199
20. von Elm E, Altman DG, Egger M, et al.: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008; 61:344–349
21. Public Health England: Oxford Health Profile. 2017.
22. Hogan H, Healey F, Neale G, et al.: Preventable deaths due to problems in care in English acute hospitals: A retrospective case record review study. *BMJ Qual Saf* 2012; 21:737–745
23. Braun, Virginia; Clarke V: Using thematic analysis in Psychology. *Qual Res Psychol* 2006; 3:77–101
24. Ke LS: Frailty in the elderly: A concept analysis. *J Nurs* 2013; 60:105–110
25. Pope C, Ziebland S, Mays N: Qualitative research in health care. Analysing qualitative

data. *BMJ* 2000; 320:114–6

26. Robert R, Coudroy R, Ragot S, et al.: Influence of ICU-bed availability on ICU admission decisions. *Ann Intensive Care* 2015; 5:1–7

27. Robert R, Reignier J, Tournoux-Facon C, et al.: Refusal of intensive care unit admission due to a full unit: Impact on mortality. *Am J Respir Crit Care Med* 2012; 185:1081–1087

EVALUATION OF A DIGITAL SYSTEM TO PREDICT UNPLANNED ADMISSIONS TO THE INTENSIVE CARE UNIT: A MIXED-METHODS APPROACH

Title Page

Authors

James Malycha MBBD FCICM[1,2,3], Oliver Redfern MBBS PhD[1], Marco Pimentel DPhil[1], Guy Ludbrook PhD MBBS FANZCA[2], Duncan Young BM DM FRCA FMedSci[1], Peter Watkinson MD ChB MRCP FDICM[1]

Affiliations

[1]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[2]Discipline of Acute Care Medicine, University of Adelaide, South Australia, Australia
[3]Intensive Care Unit, The Queen Elizabeth Hospital, Adelaide, South Australia, Australia

Keywords

Clinical deterioration, intensive care unit, critical care unit, predictive score, electronic patient record, qualitative medical note review

Wordcount

3093 words

Abstract

Background

We have developed the Hospital Alerting Via Electronic Noticeboard (HAVEN) which aims to identify hospitalised patients most at risk of reversible deterioration. HAVEN combines patients' vital-sign measurements with laboratory results, demographics, comorbidities using a machine learnt algorithm.

Objectives

The aim of this study was to identify variables or concepts that could improve HAVEN predictive performance.

Methods

This was an embedded, mixed methods study. Eligible patients with the five highest HAVEN scores in the hospital (i.e., 'HAVEN Top 5') had their medical identification details recorded. We conducted a structured medical note review on these patients 48 hours post their identifiers being recorded. Methods of constant comparison were used during data collection and to analyse patient data.

Results

The 129 patients not admitted to ICU then underwent constant comparison review, which produced three main groups. Group 1 were patients referred to specialist services (n = 37). Group 2 responded to ward-based treatment, (n = 38). Group 3 were frail and had documented treatment limitations (n = 47).

Conclusions

Digital-only validation methods code the cohort not admitted to ICU as 'falsely positive' in sensitivity analyses however this approach limits the evaluation of model performance. Our study suggested that coding for patients referred to other specialist teams, those with treatment limitations in place, along with those who are deteriorating but then respond to ward-based therapies, would give a more accurate measure of the value of the scores, especially in relation to cost-effectiveness of resource utilisation.

.

Introduction

Early Warning Score (EWS) systems have been adopted internationally to identify patients who deteriorate in acute hospitals. (1) EWSs combine individual vital-sign observations abnormalities into a single score. These scores are easily calculated at the bedside and alert clinicians to ward patients at high risk of clinical deterioration. However, EWSs do not account for additional patient risk factors, limiting their sensitivity and generating false alerts. (2) Recent studies have shown how scoring systems that use additional variables from the Electronic Patient Record (EPR) (e.g., laboratory results and comorbidities) outperform EWS systems that rely solely on vital signs. (3–10)

We developed the Hospital Alerting Via Electronic Noticeboard (HAVEN). (11) To develop the HAVEN model, we used standard statistical methods to select variables identified with a modified Delphi panel of experts and systematic review of existing literature. (12) HAVEN uses a machine learnt algorithm to combine patients' vital-sign measurements with their laboratory results, demographics, and comorbidities into a single risk score. It was developed and externally validated to predict impending cardiac arrests and unplanned transfers to the Intensive Care Unit (ICU). (13).

There are clear guidelines for developing, validating and reporting prognostic models in healthcare. (14) The guidance includes using appropriate statistical metrics (e.g., discrimination, calibration) for assessing model performance. However, rigorous methods for assessing their clinical utility and identifying factors that could improve model performance are limited. (15–17) Moreover, in contrast to EWS systems that alert when the score exceeds a threshold, HAVEN ranks all patients in the hospital from highest to lowest risk of having an adverse event.

The aim of this study was to discover additional variables, not recognised during the data-driven development process, that would improve the performance of the HAVEN risk score.(12) To do this, we reviewed the medical records of misclassified patients, that is, patients ranked highly by HAVEN but who were not admitted to the ICU; or patients who were never ranked highly by HAVEN but had an unplanned ICU admission. The general hypothesis was that identification of misclassification, such as patients false positively identified as in need of ICU care, would provide data to refine future identification and classification of 'at risk' patients.

Methods

A protocol for the study was published in advance (11). The study is reported according to the STrengthening Reporting of OBservational studies in Epidemiology (STROBE) guidelines **(Appendix, Table 1)**. (18)

Setting

The study was conducted at the John Radcliffe Hospital (JRH) which is part of the Oxford University Hospitals National Health Service Foundation Trust. The JRH is an 800-bed hospital that serves a local population of around 800,000 and a wider tertiary referral population.

Design

This was an embedded, mixed methods study. Methods of constant comparison were used during data collection and to analyse patient data extracted from the EPR. (19)(20)

Participants

Participants were identified using a web interface that displayed real-time HAVEN risk scores for all currently admitted JRH patients. (21) At 9am, on 27 out of the 39 (i.e., researcher available) days between July 27th and September 3rd, 2019, eligible patients with the five highest HAVEN scores in the hospital (i.e., 'HAVEN Top 5') had their identifiers recorded.

Inclusion and Exclusion Criteria

We included all adult patients (aged >16 years of age) admitted to the study hospital. We excluded patients on high acuity, obstetric or paediatric wards at the time of the HAVEN Top 5 being recorded. High acuity excluded wards were the: intensive care unit, coronary care unit, specialty respiratory ward, high dependency unit, obstetric and paediatric wards.

Data Collection & Analysis

Both quantitative and qualitative methods were used to collect and analyse data. Data were collected between 48 hours and 60 hours post the HAVEN alert being documented (i.e., day three). Patients in the top five highest scoring patients each day were sub-divided into True Positive or False Negative, depending on whether they were admitted to ICU in this 48-hour period (i.e., False Positive *were not* admitted to ICU, True Positive *were* admitted to ICU). Concurrently, patients admitted to ICU during designated study days without being in the five highest ranked HAVEN scores during the preceding 48-hours were categorised as False

Negative. (11) Cardiac arrest was not used to define False Positive or Negative because of its very low event rate within the hospital population.

Medical Note Review

Data collection was conducted using medical note review, by a single researcher (JM) who was a consultant in critical care medicine. The process was a modified version of the methods published by Hogan et al. (22) Data were captured using Microsoft Excel and stored in a secure local server.

Core Data

Core patient-centred and system-level data collected for the study were agreed a priori. (11) These included patient age, sex, admission date, primary diagnosis, admitting team, elective or emergency admission status, prior surgery, past medical history, treatment limitations, current medications, radiological imaging, discharge status and Clinical Frailty Scores. (23) System-level variables included admission specialities (e.g., medical or surgical) and inter-hospital transfers.

Constant Comparison

Data were analysed using methods of constant comparison (CC). (20) CC is based on Grounded Theory, first developed by Glaser & Strauss in 1968. (24) CC (also called 'Theoretical Sampling') is a qualitative research method where data collection, coding and classification iterates and evolves as the study progresses. This process continues until a replicable 'theory' is produced which optimally explains the relationships within the data, specific to a question being asked of the data. (25)

Data Coding

Data coding was conducted in three phases: Open, Axial and Selective coding. Open coding compartmentalised the data into discrete (i.e., single) blocks or concepts. Axial Coding grouped the Open Codes and informed the iterative data collection process. The Selective Code (i.e., the Theory) was developed via analyses of the Axial Codes, which in this case was the classification process that best explained the False Positive cohort in this study. The data coding process was conducted using Taguette© (an opensource software alternative to NVivo©) and Microsoft Excel.

Additional Data

As Open and Axial Coding proceeded during the data collection process, additional data fields were added.

Sample Size

Sample size was dictated by pragmatic considerations of researcher capacity and availability and is described in detail in the protocol. (11) Theoretical saturation (which is the point at which no new information or concepts are gained from ongoing data collection) was not considered when deciding sample size. (26)

Results

140 patients were identified for inclusion in the cohort with 134 patients having medical note data extracted (6 were excluded because of incomplete medical record keeping). The median HAVEN score was 82, with a variance of 44.2 **(Appendix, Table 2**). The mean, median and minimum HAVEN scores for each study day are shown in the **Appendix, Table 3**. Of the 134 patient, 129 were not admitted to ICU (i.e., False Positives) and five were admitted to ICU (i.e., True Positives). The 129 False Positive patients then underwent constant comparison **(Figure 1)**.

False Positives

The median age for the 129 False Positive cohort was 67 years and 79 were male. Forty (31%) were admitted under general medical teams, 47 (36%) under specialist medical teams, 12 (9%) under general surgical teams and 30 (23%) under specialist surgical teams (e.g., neurosurgery, cardiothoracic, ear-nose and throat, plastics, orthopaedics and trauma). Twenty-four (19%) patients had less than two known comorbidities, 63 (49%) had 2 – 4 comorbidities, 28 (22%) had 5 – 6 comorbidities and 14 (11%) had more than six comorbidities. The most common reason for hospital admission was community acquired pneumonia (17 patients, 13%). 47 patients (36%) had active Do Not Attempt Cardiopulmonary Resuscitation Orders. These data are summarised in the **Appendix (Table 4)**.

Qualitative Analysis

The Open, Axial and Selective codes produced from CC are shown in **Table 1**.

Eight Axial Codes emerged as stepwise explanatory factors for why a patient scored as high risk by HAVEN but did not require a subsequent ICU admission. *Context* grouped Open Codes

specific to unchangeable but relevant patient centred variables that act as coefficients for decisions during the patient's admission. *Patient Group* grouped Open Codes were descriptors. *Events* were defined as episodes during the patient's admission with an objectively observed and documented explanation for a transition from a more to less stable clinical state. *Interventions* were documented actions taken by treating clinicians as a response to *Events. Pathology* were recorded investigation results (arterial blood gas, laboratory, routine, non-routine (i.e. tests other than standard blood count, liver function, biochemistry and basic coagulation studies), single and multiple. In each case the data provided an objective marker of disease severity and trajectory. *Physiology* was vital-sign observation sets, grouped into baseline (normal/abnormal) and trajectories (normal → abnormal and vice versa). These data provided an objective marker of physiological response to *Pathology*. *Opinion* was documentation in the EPR medical note, stating a clinical opinion regarding *Pathology, Physiology* and overall clinical status. *Outcomes* were Open Codes that were the result *of Events, Interventions* and *Objective Data*. These included change in location, death, ICU admission, specialist referral, treatment types, limitations and response (or otherwise) to ward based therapy. *Outcomes* represented the grouped Open Codes with the most relevant and highest accumulation of objective data regarding the patient's clinical state, progress and reason for their not requiring an ICU admission. The accumulation of objective patient data is represented in **Figure 2**.

Outcomes were divided into four main groups. Group 1 were clinically unstable patients but were referred to another specialist service with higher monitoring and intervention capabilities (e.g., cardiology and respiratory wards) for ongoing management. Group 2 patients were clinically unstable but received timely ward-based treatments, such as intravenous fluids or antibiotics and did not need ICU referral or admission. Patients in Group 3 were clinically unstable but were commonly frail and/or had documented treatment limitations, where it had been agreed that transfer to ICU would not be in their best interests. Group 4 were clinically stable and had no imminent need for ICU admission. Specific descriptions of the *Outcomes* are shown in the **Appendix (Table 5)**.

Table 2 shows the distribution of False Positive patients across the four groups described above (i.e. *Outcomes*). Of the 129 patients, 37 had been referred to non-ICU specialist teams *(Group 1)*, 14 were female and the median age was 60 years. Group 1 had fewer patients with significant frailty (i.e. Frailty Scale Score of 5 or greater) (n = 11, 29%). 38 patients responded

to ward-based therapies *(Group 2)*. This group also had a low number of patients with significant frailty (n = 5, 13%), with 16 females and median ages of 65 years. 47 patients had treatment limitation *(Group 3)*. This groups had a higher proportion of significantly frail patients (n = 38, 80%) and a higher median age of 77 years. Of the seven patients who were in the FHR cohort but were objectively well *(Group 4)*, none were severely frail. Four had been discharged from acute care settings post cardiac surgery but had residual, stable derangements in physiological parameters that were documented but of low clinical concern.

Specialist team referrals are shown in the **Appendix, Table 6.** Documented reason for treatment limitations are shown in the **Appendix, Table 7.**

True Positives

Five True Positives patients were identified. All were males aged between 51 and 69. Three were general surgical (two with post-operative respiratory failure and one with a seizure), one was neurosurgical, with a subdural haematoma with respiratory failure and one was a medical patient with respiratory failure. All had Frailty Scale Scores of 1 – 4.

False Negatives

Three eligible patients were admitted to ICU during the study period and not ranked within the Top 5 by the HAVEN model. One patient had sudden, acute respiratory failure and ultimately received palliative care in the ICU, one patient developed an acute delirium in the day following a carotid endarterectomy (which did not produce any physiological derangement). The third patient became narcotised (with low respiratory rate) on the second day after a total knee replacement and required a naloxone infusion. Patient numbers in this low ranked cohort were insufficient for classification via qualitative methods.

Discussion

Key results

This study is one of very few to prospectively evaluate a machine learning algorithm using a mixed-methods approach. During the four weeks of the study, only five patients were False Negatives (i.e., admitted to the ICU ranked outside the Top 5 by HAVEN). However, 129 (96.2%) patients were False Positives (i.e., ranked in the Top 5 by HAVEN but were not admitted to an ICU). Of these False Positive patients, the majority were objectively unstable but only a small proportion required admission to ICU.

Using methods of constant comparison, we demonstrated there are four main groups of patients that explain these findings. Firstly, patients requiring referral to other specialist teams (e.g., cardiology). Secondly frail and often elderly patients with treatment limitations. Thirdly, patients who were promptly treated on the general ward and therefore avoided requiring an ICU admission. Finally, but infrequently, clinically stable patients with no need for ICU involvement.

Strengths and limitations

To our knowledge, this is the first study to apply an embedded mixed-methods approach to evaluate a prognostic risk score for deteriorating ward patients. Machine-learning in healthcare is increasingly being studied for potential use, but with the notable exception for Escobar et al., examples of successful, real-world implementation are uncommon. (27)(28) Our mixed-methods approach to evaluating HAVEN could provide a complementary approach to assessing similar clinical decision support systems. An important limitation of this study was not considering theoretical saturation when calculating the sample size. This was a pragmatic decision based on researcher availability. However, the Theory emerged between 70 and 80 medical record reviews and was supported through the subsequent data collection and analysis, so saturation did occur.

Once implemented, a key indicator of HAVEN's utility will be increased efficiency in the work of rapid response systems (RRS). Ranking patients by risk was used in this study to simulate the potential implementation of HAVEN via RRSs. This approach deviates from most current EWS systems, which direct clinical activity via thresholds or triggers. However, this method is not without its limitations. Analysis of retrospective data showed between one and two patients were admitted daily to the ICU from the general ward in the study hospital. Thus, we anticipated between three and four 'false results' from the outset but considered this acceptable given the intended aim was not to undertake a formal evaluation of system performance but to identify reasons for HAVEN error. This study was not designed to robustly evaluate the performance of HAVEN, but to understand the source of false positives, which will guide future development of the system. The size and resource availability of the Rapid Response Team, as well as hospital risk tolerance, will most likely be important contributors.

Comparison to current research

In a recently published systematic review, Hu et al. found 29 studies that reported the development and validation of EWS systems. (29) Of these, five were comparable with HAVEN in that they developed a prognostic model using continuous, EPR linked data (both vital-sign observation sets and additional data such a laboratory results and demographics) to predict unplanned ICU admission, although this was often as part of a composite outcome. (5, 8, 9, 30, 31) Three of these systems were tested in implementation studies (9, 30, 31) but none used qualitative methods to extract and evaluate why the model generated false results. We believe our study is the first of this kind and provides insights into how to improve efficiency. Other authors have described many aspects of digital, prognostic modelling for the deteriorating patient (32) and methods for incorporating treatment limitations when evaluating prognostic risk scores like HAVEN and note the importance of capturing these data electronically. (33)

Implications for future research

Our results suggest several factors that could improve interventions targeting deteriorating ward patients:

1. The one third of patients with clinical instability, higher median age and high frailty scores represent and important sub-group. These patients require different but no less important escalations in care to specialist geriatricians, or other units with specialised expertise. In HAVEN, generating a Frailty score relied on having had a previous admission. Developing the utility to estimate frailty in all patients within the system will form part of future work. Numerous frailty measures are found in the literature but to our knowledge only one validated digital frailty score has been published. (34)

2. Incorporating treatment limitations will be a requirement to improve efficiency when prognosticating for patients who would benefit from an ICU admission. The JRH did not have an electronic document capturing this data, so the HAVEN system was not able to incorporate it, however work is underway to add this information to the user interface and algorithm.

3. Why some patients respond to ward-based treatment and others do not is poorly understood. This group is important to identify because they are demonstrably salvaged by prompt intervention. All current metrics used in developing and validating early warning tools miss-classify these patients, which may lead to scores being developed that are less likely to recognise the very patients to whom clinicians should be called. (35) Our study results suggest this distinction may inform how to reduce alert fatigue and inefficiency.

4. Novel outcomes and objective criteria that guide escalations in care, may reduce the bias associated with traditional, system-dependent outcome measures (e.g., unplanned ICU admission) in the evaluation. Novel outcomes, including specialty referrals and need for specific treatments (e.g., renal replacement threshold), should be considered when deriving and validating prognostic models in future. Broadening outcomes is the subject of future work. Additionally, better recognition of patient sub-groups (e.g., sepsis) and common patterns of deterioration, may also inform model development.

5. Ranking patients according to risk of risk for deterioration is a novel approach and requires further study.

Conclusions

HAVEN correctly identified clinically unstable patients on the hospital ward but only a small proportion required ICU admission. Qualitative analysis demonstrated that whilst these patients did not require ICU, they were correctly identified as objectively clinically unstable. Traditional, digital-only validation methods code this cohort as falsely positive in sensitivity analyses however our study showed this approach was limited in evaluating model performance. Our study suggested that coding for patients referred to other specialist teams and those with treatment limitations in place, along with those who are deteriorating but then respond to ward-based therapies, would improve the performance of similar models. We conclude that validating a method to accurately recognise and code for sub-groups of deteriorating patients from within the EPR would be an important next step in this field of research.

Declarations

Supplemental Materials

The Appendix is provided to support the findings of this study.

Competing Interests

None declared.

Acknowledgments

JM would like to acknowledge the University of Adelaide, Department of Acute Medicine, who are administering his PhD and Mr Paul Coombes for his technical assistance. JM would like to thank the OUHNHSFT administration, clinical staff and patients for allowing this research to be conducted within their organisation.

Data Sharing

The data used in this research are not openly available.

Ethics

Approval has been granted for gathering the data used in this study (South Central – Oxford C, Research Ethics Committee) REC reference: 16/SC/0264, 13th June 2016) and Confidentiality Advisory Group (16/CAG/0066).

Funding

This research was funded by the Health Innovation Challenge Fund (HICF-R9-524; WT-103703/Z/14/Z), a partnership between the Department of Health and Wellcome Trust. The views expressed in this publication are those of the authors and not those of the Department of Health & Social Care or Wellcome Trust. PW is supported by the National Institute for Health Research Biomedical Research Centre, Oxford.

References

1. RCP: National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. 2017.

2. Ghosh E, Eshelman L, Yang L, et al.: Early Deterioration Indicator: Data-driven approach to detecting deterioration in general ward. *Resuscitation* 2018; 122:99–105

3. Tam V, Frost SA, Hillman KM, et al.: Using administrative data to develop a nomogram for individualising risk of unplanned admission to intensive care. *Resuscitation* 2008; 79:241–248

4. Churpek MM, Yuen TC, Winslow C, et al.: Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit Care Med* 2016; 44:368–374

5. Kipnis P, Turk BJ, Wulf DA, 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; 64:10–19

6. Bailey TC, Chen Y, Mao Y, et al.: A trial of a real-time Alert for clinical deterioration in Patients hospitalized on general medical wards. *J Hosp Med* 2013; 8:236–242

7. Churpek MM, Yuen TC, Park SY, et al.: Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards. *Crit Care Med* 2014; 42:841–848

8. Alvarez CA, Clark CA, Zhang S, et al.: Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. *BMC Med Inform Decis Mak* 2013; 13:28

9. Escobar GJ, Laguardia JC, Turk BJ, et al.: Early detection of impending physiologic deterioration among patients who are not in intensive care: Development of predictive models using data from an automated electronic medical record. *J Hosp Med* 2012; 7:388–395

10. Kang MA, Churpek MM, Zadravecz FJ, et al.: Real-time risk prediction on the wards: A feasibility study. *Crit Care Med* 2016; 44:1468–1473

11. Malycha J, Redfern OC, Ludbrook G, et al.: Testing a digital system that ranks the risk of unplanned intensive care unit admission in all ward patients: Protocol for a prospective observational cohort study. *BMJ Open* 2019; 9:1–4

12. Malycha J, Bonnici T, Clifton DA, et al.: Patient centred variables with univariate associations with unplanned ICU admission: A systematic review. *BMC Med Inform Decis Mak* 2019; 19:4–7

13. Pimentel MA, Redfern OC, Malycha J, et al.: Word Count. *SpringerReference* 2011;

14. Collins GS, Reitsma JB, Altman DG, et al.: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol* 2015; 67:1142–1151

15. Sanchez-Pinto LN, Venable LR, Fahrenbach J, et al.: Comparison of variable selection methods for clinical predictive modeling. *Int J Med Inform* 2018; 116:10–17

16. Fu L, Schwartz J, Moy A, et al.: Development and validation of early warning score system: A systematic literature review. *J Biomed Inform* 2020; 105:103410

17. Leisman DE, Harhay MO, Lederer DJ, et al.: Development and Reporting of Prediction Models: Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. *Crit Care Med* 2020; 48:623–633

18. von Elm E, Altman DG, Egger M, et al.: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008; 61:344–349

19. Barnett-Page E, Thomas J: Methods for the synthesis of qualitative research: A critical review. *BMC Med Res Methodol* 2009; 9

20. Boeije H: A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Qual Quant* 2002; 36:391–409

21. Dahella SS, Briggs JS, Coombes P, et al.: Implementing a system for the real-time risk assessment of patients considered for intensive care. 2020; 0:1–7

22. Hogan H, Zipfel R, Neuburger J, et al.: Avoidability of hospital deaths and association

with hospital-wide mortality ratios: Retrospective case record review and regression analysis. *BMJ* 2015; 351:h3239

23. Rockwood K, Song X, MacKnight C, et al.: A global clinical measure of fitness and frailty in elderly people. *Cmaj* 2005; 173:489–495

24. Glaser BG, Strauss AL, Strutzel E: The Discovery of Grounded Theory; Strategies for Qualitative Research. *Nurs Res* 1968; 17:364

25. Cho JY, Lee EH: Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences. *Qual Rep* 2014; 19:1–20

26. Aldiabat KM, Le Navenec CL: Data saturation: The mysterious step in grounded theory methodology. *Qual Rep* 2018; 23:245–261

27. Escobar GJ, Liu VX, Schuler A, et al.: Automated Identification of Adults at Risk for In-Hospital Clinical Deterioration. *N Engl J Med* 2020; 383:1951–1960

28. Wong A, Otles E, Donnelly JP, et al.: External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med* 2021; 181:1065–1070

29. Fu LH, Schwartz J, Moy A, et al.: Development and validation of early warning score system: A systematic literature review. *J Biomed Inform* 2020; 105:103410

30. Churpek MM, Carey KA, Dela Merced N, et al.: Validation of Early Warning Scores at Two Long-Term Acute Care Hospitals. *Crit Care Med* 2019; 47:e962–e965

31. O'Brien C, Goldstein BA, Shen Y, et al.: Development, Implementation, and Evaluation of an In-Hospital Optimized Early Warning Score for Patient Deterioration. *MDM Policy Pract* 2020; 5:1–10

32. Escobar GJ, Dellinger RP: Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med* 2016; 11:S5–S10

33. Granich R, Sutton Z, Kim YS, et al.: Early detection of critical illness outside the intensive care unit: Clarifying treatment plans and honoring goals of care using a supportive care team. *J Hosp Med* 2016; 11:S40–S47

34. Gilbert T, Neuburger J, Kraindler J, et al.: Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet* 2018; 391:1775–1782

35. Tarassenko L, Clifton DA, Pinsky MR, et al.: Centile-based early warning scores derived from statistical distributions of vital signs. *Resuscitation* 2011; 82:1013–1018

Appendix

Table 1. Strengthening Reporting of Observational studies in Epidemiology guidelines

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1,3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4,5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5,6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment	5,6

		(measurement). Describe comparability of assessment methods if there is more than one group	
Bias	9	Describe any efforts to address potential sources of bias	n/a
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	n/a
		(*b*) Describe any methods used to examine subgroups and interactions	6, App
		(*c*) Explain how missing data were addressed	App
		(*d*) If applicable, explain how loss to follow-up was addressed	n/a
		(*e*) Describe any sensitivity analyses	n/a
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	App
		(b) Give reasons for non-participation at each stage	App
		(c) Consider use of a flow diagram	App
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7
		(b) Indicate number of participants with missing data for each variable of interest	n/a
		(c) Summarise follow-up time (eg, average and total amount)	App
Outcome data	15*	Report numbers of outcome events or summary measures over time	6-8

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6-8
		(*b*) Report category boundaries when continuous variables were categorized	n/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n/a
Discussion			
Key results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9,10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	10

Table 2. Study number and HAVEN score

Study No	Last HAVEN Score
1	80
2	76
3	87
4	75
5	81
6	87
7	89
8	73
9	78
10	82
11	79
12	87
13	80
14	82
15	74
16	80
17	78
18	82
19	76
20	79
21	83
22	80
23	87
24	72
25	79
26	86
27	80
28	78
29	83
30	89
31	81
32	75
33	77
34	82
35	78
36	92
37	76
38	76
39	79
40	86
41	77

42	77
43	89
44	85
45	71
46	77
47	91
48	77
49	83
50	90
51	86
52	84
53	83
54	87
55	86
56	91
57	98
58	83
59	90
60	88
61	84
62	75
63	83
64	77
65	73
66	80
67	78
68	80
69	84
70	87
71	88
72	92
73	83
74	90
75	90
76	76
77	73
78	88
79	76
80	94
81	87
82	93
83	73
84	85
85	86
86	81

87	80
88	82
89	89
90	74
91	79
92	88
93	81
94	77
95	81
96	87
97	72
98	71
99	82
100	79
101	72
102	94
103	71
104	83
105	83
106	79
107	89
108	90
109	91
110	80
111	80
112	70
113	88
114	90
115	73
116	81
117	72
118	80
119	86
120	91
121	74
122	92
123	95
124	93
125	85
126	91
127	72
128	97
129	74
130	89
131	79

132	92
133	90
134	67

Individual study days are denoted by alternating shading/no shading (i.e., day 1 is the first 5 scores, day is the next 5 score, and so on)

Table 3. Mean, median and minimum HAVEN scores for each study day

Day	*Mean*	*Median*	*Minimum*
1	*79.8*	*80*	*75*
2	*81.8*	*82*	*73*
3	*80.4*	*80.4*	*74*
4	*79*	*79*	*76*
5	*80.2*	*80.2*	*72*
6	*83.2*	*83*	*78*
7	*78.6*	*78*	*75*
8	*81.8*	*79*	*76*
9	*79.8*	*77*	*71*
10	*83.6*	*83*	*77*
11	*85.2*	*86*	*83*
12	*90*	*90*	*83*
13	*78.4*	*77*	*73*
14	*81.8*	*80*	*78*
15	*88.6*	*90*	*83*
16	*81.4*	*76*	*73*
17	*84.8*	*86*	*73*
18	*81.2*	*81*	*74*
19	*81.2*	*81*	*77*
20	*78.2*	*79*	*71*
21	*80.6*	*83*	*71*
22	*85.8*	*89*	*79*
23	*80.2*	*80*	*70*
24	*82*	*81*	*72*
25	*87.8*	*92*	*74*
26	*84.6*	*89*	*72*
27	*82*	*84.5*	*67*

Table 4. False High Rank Patients: those ranked in the Top 5 daily scores by HAVEN as high risk for clinical deterioration who were not admitted to ICU in the subsequent 48 hours

Demographics	**Total**
Total	129
Age, median	67
Male (Female)	79 (50)
Treating team	
Medical General	40 (31%)
Medical Specialist	47 (36%)
Surgical General	12 (9%)
Surgical Specialist	30 (23%)
Comorbidities*	
0 - 1	24 (19%)
2 – 4	63 (49%)
5 - 6	28 (22%)
> 6	14 (11%)
Primary diagnosis	
IECOPD[#]	9 (7%)
Sepsis	14 (11%)
CAP[##]	17 (13%)
Trauma	8 (6%)
Other	81 (63%)
DNACPR	47 (36%)
Frailty Scale Score	
Score 1 - 4	71 (55%)
Score 5 - 6	37 (29%)
Score > 7	21 (16%)
Discharge status	
Alive	105 (81%)
Deceased	24 (19%)

[*]Comorbidities were counted and documented from the admission note, [#]IECOPD: Infective Exacerbation of Chronic Obstructive Pulmonary Disease, [##] Community Acquired Pneumonia

Table 5. Classifications of the groups derived via methods of constant comparison.

Group 1:	Referral to a specialist team	A documented record in the patient medical notes by clinical staff of a referral made, a clinical review by the referred specialist team, followed by a transfer of care to that team.
Group 2:	Responded to ward-based therapies	A documented record in the patient medical notes by clinical staff, of objective clinical instability (as per definition below) with a resultant ward-based treatment plan instigated, followed by sufficient clinical response as to avoid an ICU admission during the remainder of hospital admission (as per the time of the structured medical note review).
Group 3:	Treatment Limitation	A documented record of formalised plans regarding limits to escalations in care in the patient medical notes by clinical staff. To be classified as having Treatment Limitations, the patient required a 'DNACPR' (Do Not Attempt Cardiopulmonary Resuscitation) directive and also specific 'Not For Intensive Care Unit Admission' directive. The latter was usually found in the patient medical notes during the hospital admission or during the clinical review following a response the clinical deterioration.
Group 4:	Objectively well	No documented record of objective clinical instability in the patient medical notes by clinical staff and no recorded objective markers of physiological derangement.

Objective clinical instability: Defined as a documented record in the patient medical notes by medical and/or nursing staff of clinical concern and with specific reference to objective markers of physiological derangement (e.g., sustained, elevated heart rate). Ward-based therapies: Defined as any therapy deliverable on the general ward. These included intravenous fluids, anti-biotics and other medications. Intravenous drug infusions were not available for use on the ward. Ward nurse maximal performance was 15 minutely vital-sign observations and bedside monitoring, including 3-lead Electrocardiogram telemetry and real-time peripheral oxygen

saturation measurements via finger-probe. Non-invasive ventilation was not provided outside the Respiratory Ward.

Table 6. Specialist teams referred unstable ward patients

Speciality	Total
Acute Medicine	1
Cardiology	12
Gastroenterology	1
Hepatology	1
Infectious Disease	1
Interventional Radiology	1
Neurology	2
Obstetrics	1
Renal	2
Respiratory	8
Stroke	1
Surgical	5
Trauma	1

Table 7. Documented reason for Treatment Limitations in ICU admission negative patients

Reason	**Total**
Frailty Syndrome (poor physiological reserve)	20
Poor prognosis because patient not responding to therapy	2
Significant and irreversible underlying pathology (e.g., cancer, heart failure)	22
The primary diagnosis was not reversible	3

Chapter 4. Validating HAVEN

HAVEN and comparable systems have been shown to increase the precision with which deteriorating patients are detected in retrospective observational studies. However, few are externally validated (i.e., tested in data sets outside the one in which they were developed) and even fewer proceed to implementation. One hypothesis for this consistent inability to implement is the failure to consider whether they add value in clinical practice. Indeed, most current and in-use EWS systems are not optimised to identify patients with reversible deterioration (i.e., for those cases in which intervention is likely to change patient outcomes) but largely alert for patients with irreversible deterioration because of age, frailty, or chronic underlying conditions. HAVEN was externally validated to demonstrate its ability to identify patients with potentially reversible deterioration. In this study HAVEN was shown to perform better than both traditional EWS (such as NEWS) and comparable contemporary machine learning derived systems.

DETECTING DETERIORATING PATIENTS IN HOSPITAL: DEVELOPMENT AND VALIDATION OF A NOVEL SCORING SYSTEM

Authors

Marco A. F. Pimentel DPhil[1,7], Oliver Redfern MBBS PhD[1], James Malycha MBBS FCICM[1,2,3], Paul Meredith PhD[4], David Prytherch PhD CSci MIPEM[5], James S Briggs BA DPhil[5] Duncan Young BM DM FRCA FMedSci[1], David A. Clifton MEng DPhil (Oxon) CEng[6], Lionel Tarassenko CBE FREng FMedSci MA DPhil CEng FIET[6], Peter Watkinson MD ChB MRCP FDICM[1]

Affiliations

[1]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[2]Discipline of Acute Care Medicine, University of Adelaide, South Australia, Australia

[3]Intensive Care Unit, The Queen Elizabeth Hospital, Adelaide, South Australia, Australia

[4]Research and Innovation Department, Portsmouth Hospitals NHS Trust, Portsmouth, UK

[5]Centre for Healthcare Modelling & Informatics, School of Computing, University of Portsmouth, UK

[6]Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK

Keywords

Clinical deterioration, intensive care unit, critical care unit, predictive score, electronic patient record, machine learning.

Funding

This publication presents independent research funded by the Health Innovation Challenge Fund (HICF-R9-524; WT-103703/Z/ 14/Z) a parallel funding partnership between the Department of Health and Wellcome Trust. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health & Social Care or Wellcome Trust. PJW, DAC and LT are supported by the NIHR Biomedical Research Centre Oxford. MAFP and OCR are supported by Drayson Research Fellowships.

Ethics

This work received Health Research Authority, Research Ethics Committee (reference number 16/SC/0264 from the Oxford University Hospitals Trust REC, and 08/02/1394 from the Isle of Wight, Portsmouth and South East Hampshire REC), and the Confidentiality Advisory Group (16/CAG/0066) approval.

Statement of Authorship

Title of paper	Detecting deteriorating patients in hospital: development and validation of a novel scoring system
Publication status	Published
Publication details	Pimentel MA, Redfern OC, Malycha J, Meredith P, Prytherch D, Briggs J, Young JD, Clifton DA, Tarassenko L, Watkinson PJ. Detecting Deteriorating Patients in Hospital: Development and Validation of a Novel Scoring System. Am J Respir Crit Care Med. 2021 Feb 1.

Principal Author(s)

Name of Principal Author	Marco Pimentel
Contribution to paper	Created Oxford dataset, study design, data analysis and interpretation, writing the manuscript
Overall percentage (%)	40%

Name of Principal Author	Oliver Redfern
Contribution to paper	Study design, data analysis and interpretation, writing the manuscript and corresponding author
Overall percentage (%)	40%

Co-author contributions

Each author certifies:

- The candidate's stated contribution to the publication is accurate (as state above);
- Permission is granted for the candidate to include the publication in the thesis; and
- The sum of all co-author contributions is equal less the candidate's stated contribution

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr James Malycha
Contribution to paper	Study design, data analysis, clinical evaluation, evaluated and edited manuscript
Overall percentage	10%

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Paul Meredith
Contribution to paper	Assisted with study design, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor David Prytherch
Contribution to paper	Assisted with study design, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor James Briggs
Contribution to paper	Assisted with study design, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Duncan Young
Contribution to paper	Assisted with study design, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor David Clifton
Contribution to paper	Assisted with study design, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Lionel Tarassenko
Contribution to paper	Assisted with study design, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Peter Watkinson
Contribution to paper	Evaluated and edited manuscript

Abstract

Rationale

Late recognition of patient deterioration in hospital is associated with worse outcomes, including higher mortality. Despite the widespread introduction of early warning score systems and electronic health records, deterioration still goes unrecognised.

Objectives

To develop and externally validate a Hospital-wide Alerting Via Electronic Noticeboard (HAVEN) system to identify hospitalised patients at risk of reversible deterioration.

Methods

A retrospective cohort study of patients 16 years of age or above admitted to four UK hospitals. The primary outcome was cardiac arrest or unplanned admission to the intensive care unit (ICU). We used patient data (vital signs, laboratory tests, comorbidities, frailty) from one hospital to train a machine learning model (gradient boosting trees). We internally and externally validated the model and compared its performance to existing scoring systems (including NEWS, LAPS-2 and eCART).

Measurements and Main Results

We developed the HAVEN model using 230,415 patient admissions to a single hospital. We validated HAVEN on 266,295 admissions to four hospitals. HAVEN showed substantially higher discrimination (c-statistic 0.901 [95% CI 0.898-0.903]) for the primary outcome within 24 h of each measurement than other published scoring systems (which range from 0.700 [0.696-0.704] to 0.863 [0.860-0.865]). With a precision of 10%, HAVEN was able to identify 42% of cardiac arrests or unplanned ICU admissions with a lead time of up to 48 h in advance, compared to 22% by the next best system.

Conclusion

The HAVEN machine learning algorithm for early identification of in-hospital deterioration significantly outperforms other published scores such as NEWS.

Introduction

Over 60,000 patients annually deteriorate on UK hospital wards to the extent they require intensive care unit (ICU) admission (1). Late or missed recognition of deterioration is associated with worse patient outcomes, including higher mortality (2–4). Over the past twenty years, health care systems worldwide have implemented alerting systems to improve detection of patients at risk of deterioration (5–7). Most are based on abnormalities in patients' vital signs, usually by combining them into an early warning score (EWS). Clinicians are alerted when the EWS rises above a given threshold. Many hospitals also employ rapid response teams to respond to the most critically unwell patients (8). However, there is conflicting evidence that implemented EWS systems or rapid response teams improve patient outcomes (8–11).

Current EWSs were designed to be calculated easily at the bedside, when most hospitals recorded observations on paper charts. This simplicity means EWSs cannot account for trends over time, patients with chronically abnormal physiology or other indicators of deterioration (e.g., acute kidney injury). Consequently, EWSs commonly generate false alerts, risking alarm fatigue and increasing the likelihood that deteriorating patients are missed (12).
Increased uptake of electronic health records (EHRs) facilitates the development of sophisticated EWSs incorporating additional routinely collected patient data. For example, our group and others have shown that combining laboratory results with vital-sign measurements increases the precision with which deteriorating patients can be detected (13–19). Many newer risk scores exploit machine learning algorithms (13, 15, 17, 20–24). However, few are externally validated (25–27) and fewer still implemented in the EHR (23). Those that have are often subject to proprietary licences, which can limit the research community's ability to validate them (22, 23, 28, 29). Some algorithms also use data, such as detailed nursing assessments, not routinely recorded in the EHR (28). A key reason predictive machine learning models are not clinically implemented is the failure to consider whether they add value in clinical practice (15, 30, 31). Indeed, we previously argued that even current EWS systems are not optimised to identify patients with reversible deterioration; i.e., where intervention is likely to change patient outcomes (32).

In this study, we describe the development and external validation of the Hospital-wide Alerting Via Electronic Noticeboard (HAVEN) system to identify patients with potentially

reversible deterioration. HAVEN provides a real-time risk assessment, continuously updated using patients' vital signs, laboratory tests and medical histories.

Methods

Study design

A multi-centre retrospective development and external validation of a prognostic model. It is reported following the TRIPOD guidelines (33).

Setting

Patient data were collected retrospectively from two separate UK hospital groups: Portsmouth Hospitals NHS Trust (PHT) and Oxford University Hospitals NHS Trust (OUHT). Data were extracted, linked and de-identified before being made available to the research team.

PHT is a large acute district general hospital (Hospital A) with approximately 1,250 beds, providing a full range of elective and emergency medical and surgical services to a local population of around 675,000 (34). OUHT is a hospital group with approximately 1,465 beds, serving a local population of around 655,000. We included the tertiary referral centre for trauma, cardiology and neurosurgery which also provides general acute medical and surgical services (Hospital B); the specialist renal transplant and cancer referral centre (Hospital C); and the district general hospital (Hospital D). We excluded a hospital performing predominantly elective orthopaedic procedures.

Data sources

Routinely collected data stored across different clinical information systems in all four hospitals were extracted. Data included admissions' administrative information (including dates and timings for admission, discharge and/or any transfers within the hospital site), diagnoses as International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes, laboratory results (including haematology, biochemistry and microbiology), vital signs and patient demographics.

Participants

We included all patients (aged 16 or above) admitted to Hospital A from January 2012 to December 2017 and Hospitals B-D from January 2016 to December 2017.

Admissions with no recorded vital signs were excluded to ensure a minimum required dataset for score computation.

The training cohort comprised admissions to Hospital A from January 2012 to December 2015. The primary test cohort combined admissions from Hospitals A-D between January 2016 and December 2017.

Outcomes

Our primary outcome was a composite of in-hospital cardiac arrest and unplanned admission to ICU not preceded by surgery in the prior 24 hours. ICU admissions shortly after surgery were excluded as deterioration may happen during the procedure, not on the ward. Secondary outcomes were unplanned admission to ICU not preceded by surgery in the prior 24 hours and in-hospital cardiac arrest separately. We included a third secondary outcome of all unplanned admissions to ICU, to determine the effect of including unplanned ICU admissions preceded by surgery within 24 hours.

Predictors

We identified potential variables for inclusion in the model by a systematic literature search (35) and expert suggestions, followed by expert panel review. The expert panel comprised critical care nurses and doctors, alongside a statistician and senior general physician. The panel undertook a modified Delphi process to consider additional variables useful in determining patients' risk of deterioration. Consensus was reached after two discussion rounds, with a final 76 candidate variable list.

Each patient admission was represented by *static* (time-invariant), and *dynamic* (i.e., time-varying,) variables.

As *static* variables, we included the patient's age and gender at admission to the hospital. We also encoded the presence or absence of comorbidities using ICD-10 diagnosis codes. As diagnostic coding in the UK typically occurs after discharge from hospital, this information is not available electronically unless the patient has previously been admitted to the same hospital. We extracted unique diagnostic codes from previous admissions over the two years prior to the hospital admission under study. Diagnostic codes were grouped into 30 categories

according to Elixhauser (36), comprising 30 binary features encoding whether patients had common chronic diseases, such as congestive heart failure or chronic lung disease. We further calculated: smoking status (using the ICD-10 codes: F17, Z716, Z720); the Hospital Frailty Risk Score (37); and the total length of all hospitalisations in the two years prior.

As *dynamic* variables, we included commonly measured laboratory tests, vital signs and estimated inspired oxygen fraction. A variable list is in Supplementary Information section D. We designed HAVEN to re-calculate a patient's deterioration risk each time a new variable is recorded. When one time-varying variable is measured, other variables often are not. We therefore included the most recent measured value for each physiological and laboratory result variable at each time point (equivalent to a last-value-carried-forward imputation). To capture how variables change over time, we also calculated two derived features before imputation: a 24-hour variability index for physiological variables (38), (defined as the difference between the maximum and minimum values over the preceding 24 hours) and the maximum and minimum values of laboratory results recorded during the patient's admission prior to the time point (both including the current measurement).

Missing data

Distributions of variables were inspected manually. A clinical expert panel identified "biologically implausible" ranges, with values outside these ranges defined as missing.

Remaining missing values were imputed with the median (or mode for dichotomous variables) of each variable from the training set. While other methods were considered, such as multiple imputation (39), we used the median and mode to simulate the HAVEN implementation within a live clinical system.

Statistical analysis

Model development

We trained the HAVEN system by generating the set of features for each time point where a new measurement (vital sign or laboratory test) occurred. We labelled each time point as "positive" if the primary outcome occurred within 24 hours. We used a gradient boosting machine with decision trees as implemented in the XGBoost library (40). XGBoost has a number of hyperparameters (e.g., the depth of the component decision trees), modifiable to

produce the best model. One of these hyperparameters changes the relative weighting between the positive and negative classes, which can improve model performance in unbalanced data sets. To discover the optimal hyperparameters, we used a random search (500 permutations) and selected the model with the highest c-statistic (using a five-fold cross-validation procedure), using the first three years of data in the training set.

Optimal model predictions were re-calibrated on the training set's final year of data to reflect the frequency of observed outcomes using isotonic regression (41). Uncalibrated and calibrated predictions were compared using reliability plots (41).

In addition to the gradient boosting machine, we trained, optimised and validated four alternative machine learning models: a Random Forest, a Generalized Additive Model, and both L1-regularised (Lasso) and L2-regularised logistic regression models (Table SA6).

Model evaluation

We evaluated risk prediction model performance using the test set containing data from all four hospitals. In line with TRIPOD guidance, we report results for individual hospitals and for the three hospitals not used to develop HAVEN (33). We report model performance using discrimination and calibration metrics computed at both the "observation" and "patient-admission" level. We designed HAVEN to identify patients at risk of deterioration on hospital wards, rather than direct admissions from the emergency department (ED) – for this reason, scores generated from ED measurements were excluded.

At the observation level, we calculated the area under the receiver operating characteristic curve (ROC AUC) for our outcome measures occurring within the subsequent 12, 24, and 48 hour periods of each measurement (i.e., each time a measurement is recorded). The ROC AUC (c-statistic) measures discrimination, corresponding to the probability that patients who experience the outcome will be ranked above those who do not. As the outcomes are relatively rare (many more patients go home without an event than have an unplanned ICU admission or a cardiac arrest), we also computed the area under the precision-recall (PR AUC) curve, which can be informative in class-imbalanced datasets (42, 43). The precision-recall curve shows the trade-off between precision (positive predictive value) and recall (sensitivity) at each threshold. The closer the PR AUC is to 1, the greater the ability of the score/model to detect true cases

(recall) with high precision over the range of thresholds. Calibration curves for selected models were determined for outcome occurrence within 24 hours of each measurement.

The sequential nature of predictions means the total number of positive time steps (where the outcome occurs within *n* hours) does not directly correspond to the number of patients experiencing the outcome. Multiple positive time steps may be associated with a single adverse event. To assess the clinical applicability of the proposed model, we calculated the "patient-admission sensitivity" at different levels of precision (5%, 10%, 20%). These precisions correspond to evaluating 20, 10, 5 patients respectively for each true positive – also known as the number needed to evaluate (NNE) (44). For each precision level, a patient-admission was considered a false positive if they had at least one score above the threshold and no adverse event occurred. True positives were patient-admissions with at least one score above the threshold in the *n* hours prior to an adverse event. We examined the sensitivity of the model over different time prediction windows preceding the event (up to 48 hours). To further evaluate clinical utility, we performed decision curve analysis (45–48).

All 95% confidence intervals were calculated using bootstrapping (200 samples) (49) . We used the SHapley Additive exPlanations (SHAP) algorithm (50) to calculate the relative "importance" of each predictor in the final model (see Supplementary Information, Section F).

Comparison with published risk scoring systems

We compared HAVEN score performance with established early warning score (EWS) systems: the centile-based EWS (CEWS) (51), the modified EWS (MEWS) (52), the standardised EWS (SEWS) (53), the National EWS (NEWS) (54), and the cardiac arrest risk triage (CART) score (55). We also compared it with three physiological scoring systems: the NEWS:LDTEWS (13), the eCART (electronic CART) score (56), and the laboratory-based acute physiology score (LAPS-2) (57). We excluded scoring systems where the coefficients were unpublished or where data (e.g., nursing assessments) were not routinely recorded in our study sites (22, 58). Further details of EWSs/scoring systems are in Supplementary Information section C.

Results

After exclusions, we included 496,710 unique admissions to four hospitals. The training set included 230,415 admissions (from 113,450 patients) to Hospital A .
There were 266,295 admissions (159,182 patients) to four hospitals (A-D) in the test set. The two cohorts have similar patient characteristics (Table 1), both with a slightly higher proportion of female patients (of around 53%) and a median age 62-63 years.

In the test cohort, 31% of admissions to the four hospitals (A-D) were elective, with a median hospital stay of 1.36 (interquartile range, IQR, 0.36-4.76) days. Hospital mortality was ~3%. In ~1.0% of admissions, patients had an unplanned ICU admission without visiting theatre in the preceding 24 hours. A cardiac arrest occurred during 647 admissions (0.2%). There was some variability in patient characteristics across the four hospitals (see Supplementary Information Table SA1). Hospital C had a higher proportion of elective admissions (55.6%) than the other hospitals, a lower mortality rate (1.9%), and a higher rate of unplanned ICU admissions (3.9%). Class-imbalance and the extent of missing data are reported in Supplementary Information section E.

The calibration curve in the combined test set is shown in Figure 1. Table 2 shows HAVEN model performance on the test set for predicting the observation-level primary outcome (unplanned ICU admission or cardiac arrest) within different time windows. ROC AUC values increase as the time window moves closer to the event, from 0.881 (95% confidence interval, CI, of 0.879 to 0.883) within the following 48 hours, to 0.921 (95% CI of 0.919 to 0.924) within the following 12 hours. A similar trend in ROC AUC values occurs for the individual secondary outcomes (Table 2). HAVEN model performance (either by ROC or PR AUCs) was higher for predicting unplanned ICU admissions than for cardiac arrests (Table 2b). The average contributions (“feature importance”) of individual predictors are shown in Supplementary Section F.

HAVEN performance was higher than all other published EWS/risk scores when predicting the primary outcome, measured by either the ROC AUC or PR AUC (Table 3). For example, for a time window of 24 hours, HAVEN had a ROC AUC of 0.901 (95% CI 0.898 to 0.903) versus 0.863 (0.860 to 0.865) for LAPS-2, the next best performing scoring system. This improved performance remained when testing was restricted to individual hospitals (Table SA4) and to the three test hospitals (B-D) where HAVEN had not been developed (Table SA5).

HAVEN performed as well or better than all other EWS/risk scores for the individual secondary outcomes (see Supplementary Information Table SA3).

Figure 2 shows the patient-admission level sensitivity of HAVEN for different prediction time windows for three fixed levels of precision. A greater proportion of events were correctly predicted, as outcomes are included closer to the prediction point. At 10% precision (Number Needed to Evaluate, NNE, of 10), HAVEN identified 42% of adverse events occurring in the subsequent <1-48h, and 27% of adverse events occurring between 12 and 48 hours after the prediction point. In comparison, LAPS-2 identified 22% and 14% of adverse events in the corresponding time periods (Supplementary Information Figure SB1). NEWS and LAPS-2 performed similarly. The total number of events becomes smaller as the window duration decreases. Nearly all patients were in hospital for an hour prior to an event, but progressively fewer as the prediction horizon increases (roughly 60% of events occurred more than 24 hours after transfer to a general ward).

Decision curve analysis showed HAVEN had a higher net benefit than all other scoring systems over a range of risk thresholds (see Supplementary Information Figures SB3 and SB4). Including unplanned ICU admissions preceded by a theatre visit decreased the performance of HAVEN and all other scoring systems (see Supplementary Information Table SA4).

Discussion

Main findings

In this large retrospective observational study, we developed a novel risk score (HAVEN) to identify hospitalised patients at risk of potentially reversible deterioration. HAVEN had higher discrimination than all previously published EWSs and physiological scoring systems we tested (Tables 2 and 3) and was well-calibrated (Figure 1). At 10% precision, the model identified nearly twice as many adverse outcomes in advance of the event (depending on the prediction horizon, Figure 2), as the next best scoring system – LAPS-2 (Figure SB1).

Strengths and limitations

Our study used data from four large hospitals and follows the latest recommendations for developing and validating prediction models and early warning scores (45, 59). We used a composite primary outcome of unplanned admission to ICU and in-hospital cardiac arrest as a proxy for potentially reversible clinical deterioration, as no well-defined indicator of

"reversible" deterioration is recorded. This contrasts with other studies that either used only one of these two outcomes or used in-hospital mortality (60–62). We deliberately excluded in-hospital mortality from our composite outcome. In the UK, 40-50% of deaths occur in hospital and only 3.6% of these are estimated to be avoidable (63, 64). Excluding in-hospital mortality reduces the risk that our model would be optimised to predict inevitable, rather than potentially preventable, deterioration. The importance of outcome selection has been noted previously by ourselves and others (32, 61). LAPS-2 was optimised to predict in-hospital mortality, which may impact its performance in our study.

We excluded unplanned ICU admissions preceded by an operating theatre visit from the primary outcome. We assessed the impact of this exclusion on HAVEN performance, finding (as with other scoring systems) lower performance when including unplanned ICU admissions preceded by a theatre visit. This decrease was particularly marked in Hospital C, a dedicated centre for cancer and renal services (including transplants). Notwithstanding the case-mix differences in comparison to the other three hospitals (see Supplementary Information Table SA1 and Figure SB2), certain surgical procedures are undertaken on physiologically stable patients, who are routinely transferred to ICU post-operatively and coded as an unplanned ICU admission. This again demonstrates the importance of selecting the appropriate outcome when evaluating risk scoring systems.

There are limitations to using unplanned ICU admission and cardiac arrest as outcomes. These outcomes are affected by existing treatment limitation plans and Do Not Attempt CardioPulmonary Resuscitation (DNACPR) decisions. Electronic coding of these decisions varies between hospitals and is currently insufficiently robust for inclusion in a generalisable model. A recent systematic review found ICU admission can be affected by clinicians' experience, perception of benefit and organisational factors (e.g., bed availability) (65). Training our model on retrospective data risks incorporating these potential "cultural biases". We sought to reduce bias (e.g., against older patients) by including a broad range of patient factors (co-morbidities, frailty) in our model. Indeed, Figure SF3 shows that although, on average, patients aged over 80 years have a decreasing likelihood of either cardiac arrest or ICU transfer, there is wider variation in the overall predicted risk **for each age value above 80**.

To further evaluate HAVEN's predictive performance, we computed the percentage of adverse events identified ahead of time (Figure 2). We used a patient-level approach to determine the sensitivity of the model at different precision levels. As HAVEN was targeted at patients who deteriorate on general wards (rather than direct ICU admissions), we only included time periods after patients were transferred to a general ward. Our results therefore cannot be applied to patients who deteriorated in the ED. Despite the low prevalence of the outcome, the HAVEN model identified 42% of adverse events up to 48 hours in advance at a NNE of 10. Though nearly twice as good as the next best system (LAPS-2) seeing 10 patients to detect one would still create a significant workload. However, decision curve analysis (Figures SB4 and SB5) showed HAVEN has higher net benefit than the next three highest scoring systems (including NEWS, in common use in the UK). Together, these findings suggest implementing the HAVEN model should improve patient care.

Studies of early warning and other risk scores for identifying deteriorating patients vary in the outcomes and statistical methods used to validate their performance, making comparisons difficult (22, 43, 45, 66). A large retrospective study of the Advanced Alert Monitor (AAM) score, showed that AAM had a discrimination (ROC AUC) of 0.82 compared to 0.79 and 0.76 for eCART and NEWS respectively for predicting unplanned ICU admissions within 12 hours (22). In contrast, the discrimination of HAVEN was 0.92 for predicting unplanned ICU admissions *within 24 hours*, outperforming AAM over a longer prediction horizon.

Conclusion

HAVEN performed significantly better than other published scores, such as NEWS and LAPS-2, when externally validated on an independent test set. Through the use of an ensemble of "weak learners" (gradient boosting decision trees) as our machine learning algorithm, we were better able to model patients' physiological measurements in the context of their known comorbidities and frailty. We plan further external validation to ensure HAVEN model performance is sustained in other hospitals, prior to prospective evaluation on patient outcomes.

Tables

Table 1.

Summary description statistics for the cohorts. The median and interquartile range is shown for continuous variables.

		Training	Test
Patients			
Unique patients		113,450	159,182
Age, years*		63 [44-77]	62 [43-77]
Gender:	Males	52,720 (46.5%)	74,812 (47.0%)
	Females	60,730 (53.5%)	84,370 (53.0%)
Ethnicity:	White	93,853 (82.7%)	120,706 (75.8%)
	Mixed	337 (0.3%)	883 (0.6%)
	Black	437 (0.4%)	1,196 (0.8%)
	Asian	593 (0.5%)	2,698 (1.7%)
	Other	543 (0.5%)	1,280 (0.8%)
	Unknown	17,687 (15.6%)	32,421 (20.4%)
Admissions within period			
Hospital sites		1**	4**
Period		Jan 2012 – Dec 2015	Jan 2016 – Dec 2017
Unique Admissions		230,415	266,295
- per patient:	Median	1 [1-2]	1 [1-2]
	Average	2.03 (2.55)	1.67 (1.66)
Length of stay, days		1.77 [0.54-5.26]	1.36 [0.36-4.76]
Elective admissions		81,703 (35.5%)	82,402 (30.9%)
Surgical admissions		102,603 (44.5%)	116,459 (43.7%)
In-hospital deaths		7,436 (3.2%)	7,880 (3.0%)
ICU admissions***	Unpl.	2,863 (1.2%)	4,098 (1.5%)
	Unpl. Med.	2,004 (0.9%)	2,527 (0.9%)

Cardiac arrests	808 (0.4%)	647 (0.2%)
Primary outcome[#]	2,695 (1.2%)	3,105 (1.2%)

* When multiple admissions are present, the age of the patient at the first admission is used.

** In total, data from four hospitals were included: three hospitals from one organisation (Oxford University Hospitals NHS Foundation Trust) and one hospital from a different organisation (Portsmouth Hospitals NHS Trust). Data from the latter (collected within different periods), was used for training and calibration.

*** Unpl.: admissions to the ICU defined as unplanned (or unanticipated); Unpl. Med.: admissions to ICU defined as unplanned and not preceded by a visit to the theatre in the preceding 24 hours.

[#] Primary (composite) outcome is defined as occurrence of a cardiac arrest and/or an unplanned admission to the ICU.

Table 2.

Model performance using the entire test set. (a) ROC AUC and (b) PR AUC performance when predicting the risk of future adverse event (and each event separately; i.e., unplanned admission to ICU and cardiac arrest) across different time windows.

a

	ROC AUC [95% CI]		
Time window	Composite outcome[#]	Unplanned ICU admission	Cardiac arrest
12h	0.921 [0.919-0.924]	0.939 [0.936-0.941]	0.831 [0.823-0.840]
24h	0.901 [0.898-0.903]	0.921 [0.919-0.923]	0.807 [0.800-0.814]
48h	0.881 [0.879-0.883]	0.902 [0.900-0.904]	0.772 [0.765-0.779]

b

	PR AUC [95% CI]		
Time window	Composite outcome[#]	Unplanned ICU admission	Cardiac arrest
12h	0.073 [0.069-0.078]	0.076 [0.071-0.081]	0.006 [0.003-0.010]
24h	0.080 [0.076-0.084]	0.083 [0.079-0.087]	0.006 [0.003-0.008]
48h	0.081 [0.078-0.084]	0.084 [0.080-0.087]	0.006 [0.003-0.008]

[#] Composite outcome is defined as occurrence of a cardiac arrest and/or an unplanned admission to the ICU.

AUC: area under the curve; ROC: receiver operating characteristic; PR: precision-recall; CI: confidence interval.

Table 3.

Comparative performance of scoring systems using the entire test set. ROC AUC and PR AUC performance when predicting the risk of future composite adverse event (unplanned admission to ICU and cardiac arrest) within 24 hours.

Scoring systems	ROC AUC [95% CI]	PR AUC [95% CI]
CEWS	0.838 [0.834-0.841]	0.031 [0.028-0.033] ****
MEWS	0.836 [0.833-0.839]	0.031 [0.028-0.033]
NEWS	0.842 [0.839-0.845]	0.028 [0.025-0.030]
SEWS	0.791 [0.788-0.795]	0.026 [0.024-0.028]
NEWS:LDTEWS	0.860 [0.858-0.863] *	0.029 [0.026-0.031]
CART	0.700 [0.696-0.704]	0.023 [0.021-0.025]
eCART	0.796 [0.792-0.800]	0.026 [0.024-0.029]
LAPS-2	0.863 [0.860-0.865] *	0.031 [0.028-0.033] ****
HAVEN	0.901 [0.898-0.903] *	0.080 [0.076-0.084] ****

* Top-3 performing systems according to the ROC AUC.

** Top-3 performing systems according to the PR AUC.

AUC: area under the curve; ROC: receiver operating characteristic; PR: precision-recall; CI: confidence interval.

Figures

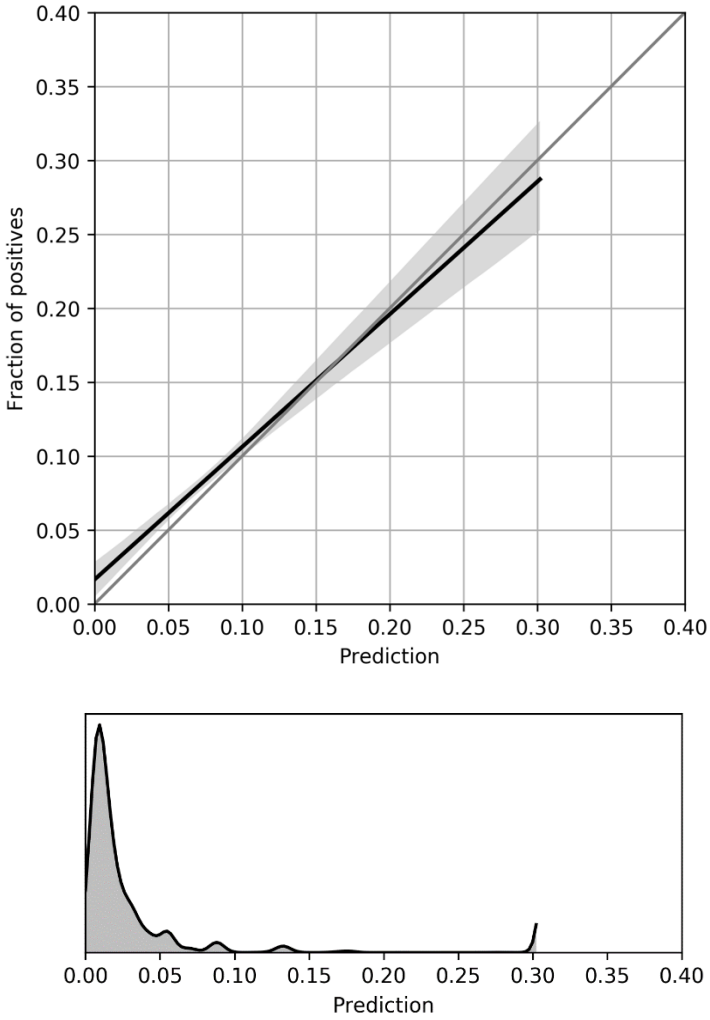


Figure 1.

Calibration curve for HAVEN predictions in the test set (top), alongside the distribution of HAVEN predictions (bottom). The calibration curve (black) shows the LOESS (locally estimated scatterplot smoothing) smoothed observed probability versus estimated probability of adverse events (with 95% confidence bands). The diagonal line (grey) shows ideal calibration.

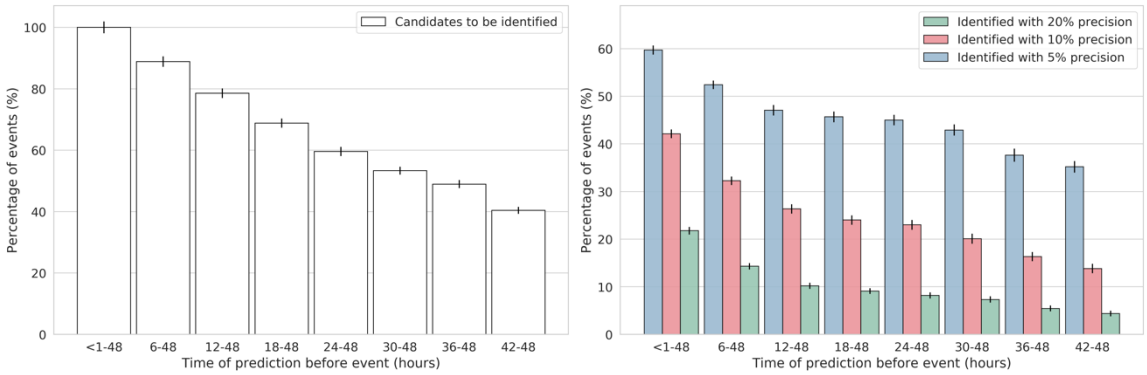


Figure 2.

Patient-admission level sensitivity: (left) average proportion of (candidate) adverse events to be identified within each window; and (right) average proportion of adverse events identified ahead of time for HAVEN at different precision levels (5%, 10% and 20%). Error bars denote one standard deviation.

References

1. Intensive Care National Audit and Research Centre (ICNARC). *Hospital length of stay pre-ICU admission and outcomes*. 2019.
2. Barwise A, Thongprayoon C, Gajic O, Jensen J, Herasevich V, Pickering BW. Delayed Rapid Response Team Activation Is Associated With Increased Hospital Mortality, Morbidity, and Length of Stay in a Tertiary Care Institution. *Crit Care Med* 2016;44:54–63.
3. National Confidential Enquiry into Patient Outcomes and Death. *Time to Intervene? A review of patients who underwent cardiopulmonary resuscitation as a result of an in-hospital cardiorespiratory arrest.* 2012.
4. Keogh B. Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report. *NHS* 2013;
5. Gerry S, Birks J, Bonnici T, Watkinson PJ, Kirtley S, Collins GS. Early warning scores for detecting deterioration in adult hospital patients: a systematic review protocol. *BMJ Open* 2017;7:e019268.
6. Smith GB, Prytherch DR, Jarvis S, Kovacs C, Meredith P, Schmidt PE, Briggs J. A Comparison of the Ability of the Physiologic Components of Medical Emergency Team Criteria and the U.K. National Early Warning Score to Discriminate Patients at Risk of a Range of Adverse Clinical Outcomes*. *Crit Care Med* 2016;44:2171–2181.
7. Pedersen NE, Rasmussen LS, Petersen JA, Gerds TA, Østergaard D, Lippert A. A critical assessment of early warning score records in 168,000 patients. *J Clin Monit Comput* 2017;32:109–116.
8. National Institute for Health and Care Excellence (NICE). *Emergency and acute medical care in over 16s: Service Delivery and Organisation*. 2018.
9. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and Early Warning Systems (EWS) for the prevention of Intensive Care admission and death of critically ill adult patients on general hospital wards. *Cochrane Database Syst Rev* 2007;CD005529.doi:10.1002/14651858.CD005529.pub2.
10. Solomon RS, Corwin GS, Barclay DC, Quddusi SF, Dannenberg MD. Effectiveness of rapid response teams on rates of in-hospital cardiopulmonary arrest and mortality: A systematic review and meta-analysis. *J Hosp Med* 2016;11:438–445.
11. McNeill G, Bryden D. Do either early warning systems or emergency response teams improve hospital patient survival? A systematic review. *Resuscitation* 2013;84:1652–

1667.

12. Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal Impact of Implemented Early Warning Score and Best Practice Alert for Patient Deterioration. *Crit Care Med* 2019;47:49–55.
13. Redfern OC, Pimentel MAF, Prytherch D, Meredith P, Clifton DA, Tarassenko L, Smith GB, Watkinson PJ. Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: Development and validation of a multivariable model. *Resuscitation* 2018;133:75–81.
14. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit Care Med* 2016;44:368–374.
15. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, Jung K, Heller K, Kale D, Saeed M, Ossorio PN, Thadaney-Israni S, Goldenberg A. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med* 2019;25:1337–1340.
16. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
17. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019;25:30–36.
18. Kang MA, Churpek MM, Zadravecz FJ, Adhikari R, Twu NM, Edelson DP. Real-Time Risk Prediction on the Wards. *Crit Care Med* 2016;44:1468–1473.
19. Dummett BA, Adams C, Scruth E, Liu V, Guo M, Escobar GJ. Incorporating an Early Detection System Into Routine Clinical Practice in Two Community Hospitals. *J Hosp Med* 2016;11:S25–S31.
20. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data. *Ann Am Thorac Soc* 2018;15:846–853.
21. Arnold J, Davis A, Fischhoff B, Yecies E, Grace J, Klobuka A, Mohan D, Hanmer J. Comparing the predictive ability of a commercial artificial intelligence early warning system with physician judgement for clinical deterioration in hospitalised general internal medicine patients: a prospective observational study. *BMJ Open* 2019;9:e032187.
22. Kipnis P, Turk BJ, Wulf DA, LaGuardia JC, Liu V, Churpek MM, Romero-Brufau S,

Escobar GJ. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform* 2016;64:10–19.

23. Sendak MP, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, Ratliff W, Balu S. A Path for Translation of Machine Learning Products into Healthcare Delivery. *EMJ Innov* 2020;doi:10.33590/emjinnov/19-00172.
24. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, Mottram A, Meyer C, Ravuri S, Protsyuk I, Connell A, Hughes CO, Karthikesalingam A, Cornebise J, Montgomery H, Rees G, Laing C, Baker CR, Peterson K, Reeves R, Hassabis D, King D, Suleyman M, Back T, Nielson C, Ledsam JR, Mohamed S. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572:116–119.
25. O'Brien C, Goldstein BA, Shen Y, Phelan M, Lambert C, Bedoya AD, Steorts RC. Development, Implementation, and Evaluation of an In-Hospital Optimized Early Warning Score for Patient Deterioration. *MDM Policy Pract* 2020;5:238146831989966.
26. Rothman MJ. The Emperor Has No Clothes. *Crit Care Med* 2019;47:129–130.
27. Paulson SS, Dummett BA, Green J, Scruth E, Reyes V, Escobar GJ. What Do We Do After the Pilot Is Done? Implementation of a Hospital Early Warning System at Scale. *Jt Comm J Qual Patient Saf* 2020;46:207–216.
28. Finlay GD, Rothman MJ, Smith RA. Measuring the modified early warning score and the Rothman Index: Advantages of utilizing the electronic medical record in an early warning system. *J Hosp Med* 2014;9:116–119.
29. Bartkowiak B, Snyder AM, Benjamin A, Schneider A, Twu NM, Churpek MM, Roggin KK, Edelson DP. Validating the Electronic Cardiac Arrest Risk Triage (eCART) Score for Risk Stratification of Surgical Inpatients in the Postoperative Setting. *Ann Surg* 2019;269:1059–1063.
30. Lindsell CJ, Stead WW, Johnson KB. Action-Informed Artificial Intelligence—Matching the Algorithm to the Problem. *JAMA* 2020;1–2.doi:10.1001/jama.2020.5035.
31. Adibi A, Sadatsafavi M, Ioannidis JPA. Validation and Utility Testing of Clinical Prediction Models. *JAMA* 2020;324:235.
32. Tarassenko L, Clifton DA, Pinsky MR, Hravnak MT, Woods JR, Watkinson PJ. Centile-based early warning scores derived from statistical distributions of vital signs. *Resuscitation* 2011;82:1013–1018.

33. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;162:W1.
34. Care Quality Commission. Oxford University Hospitals NHS Foundation Trust. at <https://www.cqc.org.uk/provider/RTH/reports>.
35. Malycha J, Bonnici T, Clifton DA, Ludbrook G, Young JD, Watkinson PJ. Patient centred variables with univariate associations with unplanned ICU admission: a systematic review. *BMC Med Inform Decis Mak* 2019;19:98.
36. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data. *Med Care* 1998;36:8–27.
37. Gilbert T, Neuburger J, Kraindler J, Keeble E, Smith P, Ariti C, Arora S, Street A, Parker S, Roberts HC, Bardsley M, Conroy S. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet* 2018;391:1775–1782.
38. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation* 2016;102:1–5.
39. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393–b2393.
40. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T, Li M, Xie J, Lin M, Geng Y, Li Y. xgboost: Extreme Gradient Boosting. 2019;
41. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. *Proc 22nd Int Conf Mach Learn - ICML '05* New York, New York, USA: ACM Press; 2005. p. 625–632.doi:10.1145/1102351.1102430.
42. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. In: Brock G, editor. *PLoS One* 2015;10:e0118432.
43. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Meltzer DO, Gibbons RD, Edelson DP. Multicenter Development and Validation of a Risk Stratification Tool for Ward Patients. *Am J Respir Crit Care Med* 2014;190:649–655.
44. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care*

2015;19:285.

45. Gerry S, Bonnici T, Birks J, Kirtley S, Virdee PS, Watkinson PJ, Collins GS. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. *BMJ* 2020;m1501.doi:10.1136/bmj.m1501.
46. Zhang Z, Rousson V, Lee W-C, Ferdynus C, Chen M, Qian X, Guo Y. Decision curve analysis: a technical note. *Ann Transl Med* 2018;6:308–308.
47. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ, Steyerberg EW. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol* 2018;74:796–804.
48. Brown M. rmda: Risk Model Decision Analysis. 2018;
49. DiCiccio TJ, Efron B. Bootstrap confidence intervals. *Stat Sci* 1996;doi:10.1214/ss/1032280214.
50. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. 2017;
51. Watkinson PJ, Pimentel MAF, Clifton DA, Tarassenko L. Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data. *Resuscitation* 2018;129:55–60.
52. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The Value of Modified Early Warning Score (MEWS) in Surgical In-Patients: A Prospective Observational Study. *Ann R Coll Surg Engl* 2006;88:571–575.
53. Paterson R, MacLeod D, Thetford D, Beattie A, Graham C, Lam S, Bell D. Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit. *Clin Med (Northfield Il)* 2006;6:281–284.
54. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. 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;84:465–470.
55. Churpek MM, Yuen TC, Park SY, Meltzer DO, Hall JB, Edelson DP. Derivation of a cardiac arrest prediction model using ward vital signs. *Crit Care Med* 2012;40:2102–2108.
56. Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP, Park S EDCMYT, Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using Electronic Health Record Data to Develop and Validate a Prediction Model for Adverse Outcomes in the Wards*. *Crit Care Med* 2014;42:841–848.
57. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P. Risk-adjusting Hospital

Mortality Using a Comprehensive Electronic Record in an Integrated Health Care Delivery System. *Med Care* 2013;51:446–453.

58. Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous measure of patient condition using the Electronic Medical Record. *J Biomed Inform* 2013;46:837–848.
59. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol* 2015;67:1142–1151.
60. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS—Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* 2010;81:932–937.
61. Churpek MM, Yuen TC, Edelson DP. Predicting clinical deterioration in the hospital: The impact of outcome selection. *Resuscitation* 2013;84:564–568.
62. Sandroni C, D'Arrigo S, Antonelli M. Rapid response systems: are they really effective? *Crit Care* 2015;19:104.
63. Hogan H, Zipfel R, Neuburger J, Hutchings A, Darzi A, Black N. Avoidability of hospital deaths and association with hospital-wide mortality ratios: retrospective case record review and regression analysis. *BMJ* 2015;351:h3239.
64. Chukwusa E, Verne J, Polato G, Taylor R, J Higginson I, Gao W. Urban and rural differences in geographical accessibility to inpatient palliative and end-of-life (PEoLC) facilities and place of death: a national population-based study in England, UK. *Int J Health Geogr* 2019;18:8.
65. Gopalan PD, Pershad S. Decision-making in ICU – A systematic review of factors considered important by ICU clinician decision makers with regard to ICU triage decisions. *J Crit Care* 2019;50:99–110.
66. Green M, Lander H, Snyder A, Hudson P, Churpek M, Edelson D. 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;123:86–91.

Supplementary Information

Section A – Extended Results Tables

Table SA1. Description statistics for the cohorts included in the test set, per hospital site. The median and interquartile range are shown for continuous variables. Admissions included occurred between January 2016 and December 2017.

Table SA2. Model performance using the test set, per hospital site. (a) ROC AUC and (b) PR AUC performance when predicting the risk of future adverse event (using the composite outcome) across different time windows.

Table SA3. Comparative performance of scoring systems using the entire test set. ROC AUC and PR AUC performance when predicting the risk of an unplanned admission to ICU or the occurrence of a cardiac arrest within 24 hours.

Table SA4. Comparative performance of scoring systems per hospital site (using the test set) for predicting (1) primary adverse events that are not preceded by a theatre visit and (2) primary adverse events that are preceded by a theatre visit. ROC AUC [95% CI] performance is shown for predicting adverse events (unplanned admission to ICU or the occurrence of a cardiac arrest) within 24 hours.

Table SA5. Comparative performance of scoring systems in the 3 external test set hospitals (B, C, D) combined. ROC AUC performance when predicting the risk of the composite adverse event (unplanned admission to ICU and cardiac arrest), unplanned admission to ICU and the occurrence of a cardiac arrest within 24 hours.

Table SA6. Comparison of the performance of different models using the entire test set. ROC AUC and PR AUC performance when predicting the risk of future composite adverse event (unplanned admission to ICU and cardiac arrest) within 24 hours. Gradient boosting trees outperforms the Logistic regression, Generalized Additive Model and Random forest baselines in terms of both ROC and PR AUCs.

Table SA1.

	Hospital A[##]	Hospital B	Hospital C	Hospital D	All 4 hospitals
Patients					
Unique patients	62,672	71,280	15,121	18,509	159,182
Age, years*	66 [44-79]	58 [38-75]	63 [50-74]	60 [43-77]	62 [43-77]
Gender Males	28,800 (46.0%)	33,810 (47.4%)	8,283 (54.8%)	8,260 (44.6%)	74,812 (47.0%)
Females	33,872 (54.0%)	37,470 (52.6%)	6,838 (45.2%)	10,249 (55.4%)	84,370 (53.0%)
Ethnicity					
White	47,914 (76.5%)	53,357 (74.9%)	10,896 (72.1%)	15,637 (84.5%)	120,706 (75.8%)
Mixed	202 (0.3%)	521 (0.7%)	109 (0.7%)	127 (0.7%)	883 (0.6%)
Black	236 (0.6%)	762 (1.1%)	179 (1.2%)	108 (0.6%)	1,196 (0.8%)
Asian	312 (0.5%)	1,900 (2.7%)	379 (2.5%)	286 (1.5%)	2,698 (1.7%)
Other	354 (0.6%)	740 (1.0%)	127 (0.8%)	115 (0.6%)	1,280 (0.8%)
Unknown	12,438 (21.4%)	14,000 (19.6%)	3,431 (22.7%)	2,236 (12.1%)	32,421 (20.4%)
Admissions					
Unique Admissions	107,981	104,977	25,496	27,841	266,295
- per patient, Median	1 [1-2]	1 [1-2]	1 [1-2]	1 [1-1]	1 [1-2]
Average	1.72 (1.89)	1.47 (1.20)	1.69 (1.54)	1.50 (1.39	1.67 (1.66)
Length of stay, days	2.07 [0.67-6.12]	1.11 [0.31-3.81]	1.85 [0.46-5.13]	0.43 [0.24-2.37]	1.36 [0.36-4.76]
Elective admissions	34,905 (32.3%)	23,925 (22.8%)	14,164 (55.6%)	9,408 (33.8%)	82,402 (30.9%)

Surgical admissions	44,535 (41.2%)	50,364 (48.0%)	13,120 (51.5%)	8440 (30.3%)	116,459 (43.7%)
In-hospital deaths	3,892 (3.6%)	2,677 (2.6%)	490 (1.9%)	821 (2.9%)	7,880 (3.0%)
ICU admissions*** U.	1,524 (1.4%)	1,515 (1.4%)	991 (3.9%)	68 (0.2%)	4,098 (1.5%)
U. M.	1,032 (1.0%)	651 (0.6%)	829 (3.3%)	15 (0.2%)	2,527 (0.9%)
Cardiac arrests	372 (0.3%)	205 (0.2%)	25 (0.1%)	45 (0.2%)	647 (0.2%)
Primary outcome[#]	1,345 (1.2%)	852 (0.8%)	848 (3.3%)	60 (0.2%)	3.105 (1.2%)

* When multiple admissions are present, the age of the patient at the first admission is used.

** In total, data from four hospitals were included: three hospitals from one organisation (Oxford University Hospitals NHS Foundation Trust) and one hospital from a different organisation (Portsmouth Hospitals NHS Trust). Data from the latter (collected within different periods), was used for training and calibration.

*** U.: admissions to the ICU defined as unplanned (or unanticipated); U. M.: admissions to ICU defined as unplanned and not preceded by a visit to the theatre in the preceding 24 hours.

[#] Primary (composite) outcome is defined as occurrence of a cardiac arrest and/or an unplanned admission to the ICU.

[##] Same hospital site that was used for development.

Table SA2.

a

	ROC AUC [95% CI]							
Window	Hospital A		Hospital B		Hospital C		Hospital D	
12h	0.911	[0.907-0.916]	0.921	[0.916-0.925]	0.931	[0.927-0.935]	0.914	[0.900-0.927]
24h	0.890	[0.886-0.894]	0.906	[0.902-0.910]	0.903	[0.899-0.907]	0.903	[0.892-0.914]
48h	0.868	[0.865-0.872]	0.892	[0.888-0.895]	0.879	[0.875-0.883]	0.888	[0.880-0.897]

b

	PR AUC [95% CI]							
Windows	Hospital A		Hospital B		Hospital C		Hospital D	
12h	0.057	[0.050-0.063]	0.099	[0.090-0.107]	0.079	[0.069-0.088]	0.048	[0.024-0.072]
24h	0.064	[0.059-0.070]	0.104	[0.096-0.111]	0.089	[0.081-0.096]	0.060	[0.041-0.080]
48h	0.065	[0.061-0.070]	0.101	[0.095-0.107]	0.100	[0.093-0.104]	0.054	[0.039-0.068]

AUC: area under the curve; ROC: receiver operating characteristic; PR: precision-recall; CI: confidence interval.

Table SA3.

Scoring systems	Unplanned ICU admissions		Cardiac Arrests	
	ROC AUC [95% CI]	PR AUC [95% CI]	ROC AUC [95% CI]	PR AUC [95% CI]
CEWS	0.860 [0.857-0.863]	**0.030 [0.027-0.032]**	0.732 [0.723-0.740]	0.002 [0.001-0.004]
LEWS	0.857 [0.854-0.860]	0.029 [0.026-0.032]	0.737 [0.728-0.745]	0.003 [0.001-0.004]
NEWS	0.863 [0.859-0.866]	0.026 [0.023-0.028]	0.746 [0.737-0.755]	0.003 [0.001-0.004]
SEWS	0.810 [0.806-0.814]	0.024 [0.022-0.027]	0.704 [0.696-0.713]	0.002 [0.001-0.004]
NEWS:LDTEWS	**0.875 [0.872-0.877]**	0.026 [0.024-0.029]	**0.793 [0.785-0.801]**	**0.003 [0.001-0.005]**
CART	0.704 [0.699-0.709]	0.021 [0.019-0.023]	0.680 [0.671-0.688]	0.003 [0.001-0.004]
eCART	0.819 [0.814-0.824]	0.024 [0.021-0.028]	0.705 [0.700-0.709]	0.002 [0.001-0.004]
LAPS-2	**0.876 [0.873-0.879]**	**0.029 [0.026-0.031]**	**0.799 [0.792-0.806]**	**0.003 [0.001-0.005]**
HAVEN	**0.921 [0.919-0.923]**	**0.083 [0.079-0.087]**	**0.807 [0.800-0.814]**	**0.006 [0.003-0.008]**

Top-3 performance scores (for each metric and each outcome) are shown in **bold**.

AUC: area under the curve; ROC: receiver operating characteristic; PR: precision-recall; CI: confidence interval.

Table SA4.

Scoring systems	Hospital A				Hospital B			
	Adverse Event (1)		Adverse Event (2)		Adverse Event (1)		Adverse Event (2)	
CEWS	0.812	[0.807-0.818]	0.782	[0.777-0.787]	0.841	[0.835-0.846]	0.791	[0.785-0.796]
LEWS	0.817	[0.812-0.823]	0.787	[0.782-0.791]	0.839	[0.834-0.845]	0.790	[0.785-0.795]
NEWS	0.828	[0.823-0.833]	0.797	[0.793-0.802]	0.837	[0.832-0.843]	0.785	[0.780-0.791]
SEWS	0.777	[0.771-0.782]	0.751	[0.746-0.756]	0.798	[0.792-0.804]	0.751	[0.746-0.757]
NEWS:LDTEWS	**0.852**	**[0.847-0.857]**	**0.819**	**[0.814-0.824]**	**0.855**	**[0.850-0.860]**	**0.800**	**[0.795-0.805]**
CART	0.674	[0.667-0.681]	0.648	[0.642-0.654]	0.731	[0.723-0.738]	0.686	[0.680-0.692]
eCART	0.779	[0.771-0.787]	0.742	[0.737-0.747]	0.802	[0.794-0.809]	0.753	[0.747-0.760]
LAPS-2	**0.854**	**[0.850-0.859]**	**0.819**	**[0.814-0.823]**	**0.857**	**[0.853-0.862]**	**0.809**	**[0.804-0.814]**
HAVEN	**0.890**	**[0.886-0.894]**	**0.863**	**[0.859-0.867]**	**0.906**	**[0.902-0.910]**	**0.859**	**[0.855-0.864]**

Scoring systems	Hospital C				Hospital D			
	Adverse Event (1)		Adverse Event (2)		Adverse Event (1)		Adverse Event (2)	
CEWS	0.871	[0.866-0.876]	0.791	[0.785-0.796]	0.805	[0.785-0.825]	0.784	[0.765-0.803]
LEWS	0.863	[0.858-0.868]	0.780	[0.775-0.786]	0.801	[0.780-0.822]	0.781	[0.762-0.800]
NEWS	0.875	[0.870-0.880]	0.784	[0.778-0.790]	0.798	[0.777-0.820]	0.776	[0.756-0.796]

	0.800	[0.793-	0.739	[0.733-	0.791	[0.770-	0.774	[0.755-
SEWS	0.807]		0.745]		0.811]		0.793]	
NEWS:LDTE	**0.882**	**[0.877-**	**0.797**	**[0.791-**	**0.839**	**[0.821-**	**0.814**	**[0.796-**
WS	**0.886]**		**0.802]**		**0.857]**		**0.831]**	
	0.727	[0.718-	0.668	[0.661-	0.718	[0.693-	0.705	[0.683-
CART	0.735]		0.675]		0.742]		0.727]	
	0.843	[0.834-	0.751	[0.744-	0.799	[0.792-	0.778	[0.769-
eCART	0.852]		0.758]		0.805]		0.787]	
	0.876	**[0.872-**	**0.792**	**[0.787-**	**0.851**	**[0.836-**	**0.835**	**[0.820-**
LAPS-2	**0.880]**		**0.798]**		**0.865]**		**0.849]**	
	0.903	**[0.899-**	**0.817**	**[0.811-**	**0.903**	**[0.892-**	**0.883**	**[0.871-**
HAVEN	**0.907]**		**0.822]**		**0.914]**		**0.895]**	

Adverse Event (1) corresponds to the primary outcome of the study: it does not consider events preceded by a theatre visit within the previous 24 hours as positive outcomes.

Adverse Event (2) includes all adverse events as positive outcomes, even when preceded by a theatre visit.

Top-3 performance scores (for each column) are shown in **bold**.

Table SA5.

	Adverse Event (1)	Unplanned ICU	Cardiac Arrest
CEWS	0.851 (0.847-0.855)	0.867 (0.863-0.870)	0.733 (0.719-0.746)
LEWS	0.845 (0.841-0.848)	0.859 (0.855-0.863)	0.739 (0.726-0.753)
NEWS	0.847 (0.844-0.851)	0.861 (0.857-0.864)	0.748 (0.735-0.762)
SEWS	0.796 (0.792-0.801)	0.806 (0.802-0.811)	0.721 (0.707-0.735)
NEWS:LDTEWS	0.862 (0.859-0.866)	0.871 (0.868-0.875)	0.796 (0.784-0.808)
CART	0.719 (0.713-0.724)	0.720 (0.714-0.726)	0.708 (0.695-0.721)
eCART	0.802 (0.796-0.808)	0.839 (0.832-0.845)	0.753 (0.748-0.758)
LAPS2	0.866 (0.863-0.869)	0.867 (0.864-0.870)	0.794 (0.783-0.804)
HAVEN	0.906 (0.903-0.909)	0.920 (0.918-0.923)	0.800 (0.789-0.811)

Table SA6.

Scoring systems	ROC AUC [95% CI]	PR AUC [95% CI]
Gradient Boosting Trees	0.901 [0.898-0.903]	0.080 [0.076-0.084]
Random Forest	0.884 [0.882-0.886]	0.065 [0.062-0.068]
Generalized Additive Model	0.871 [0.868-0.873]	0.054 [0.051-0.057]
Logistic Regression (L1-regularised)	0.857 [0.854-0.859]	0.048 [0.045-0.051]
Logistic Regression (L2-regularised)	0.852 [0.850-0.855]	0.043 [0.040-0.046]

AUC: area under the curve; ROC: receiver operating characteristic; PR: precision-recall; CI: confidence interval.

Using the same training set and input features, we trained and optimised (using the same cross-validation approach) four other machine learning models: a Random Forest, a Generalized Additive Model, and both L1-regularised (Lasso) and L2-regularised logistic regression models. Table SA6 shows the performance of these alternative models alongside the gradient boosting machine.

In addition to optimising the (hyper-)parameters of each of the four other machine models, we explored over-sampling the minority class (i.e. positive observations). While the performance of the GBM benefited from heavy overweighting of positive observations, equivalent oversampling did not improve model performance for the other machine learning models.

Section B – Extended Results Figures

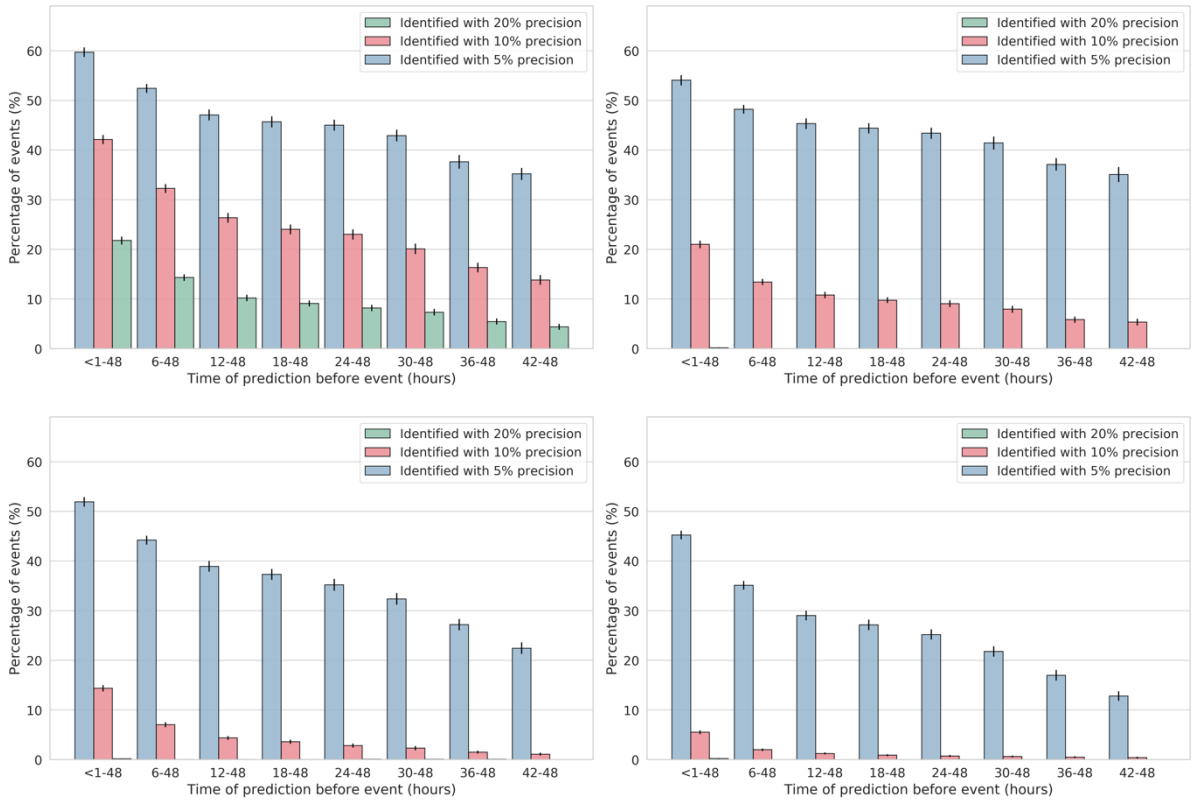


Figure SB1.

Patient-admission level sensitivity: average percentage of adverse event identified ahead of time for HAVEN (top-left), LAPS-2 (top-right), NEWS:LDTEWS (bottom-left) and NEWS (bottom-right) at different precision levels (5%, 10% and 20%). Error bars denote one standard deviation. Note: the sensitivity with a 20% precision for LAPS-2, NEWS:LDTEWS and NEWS is < 1%.

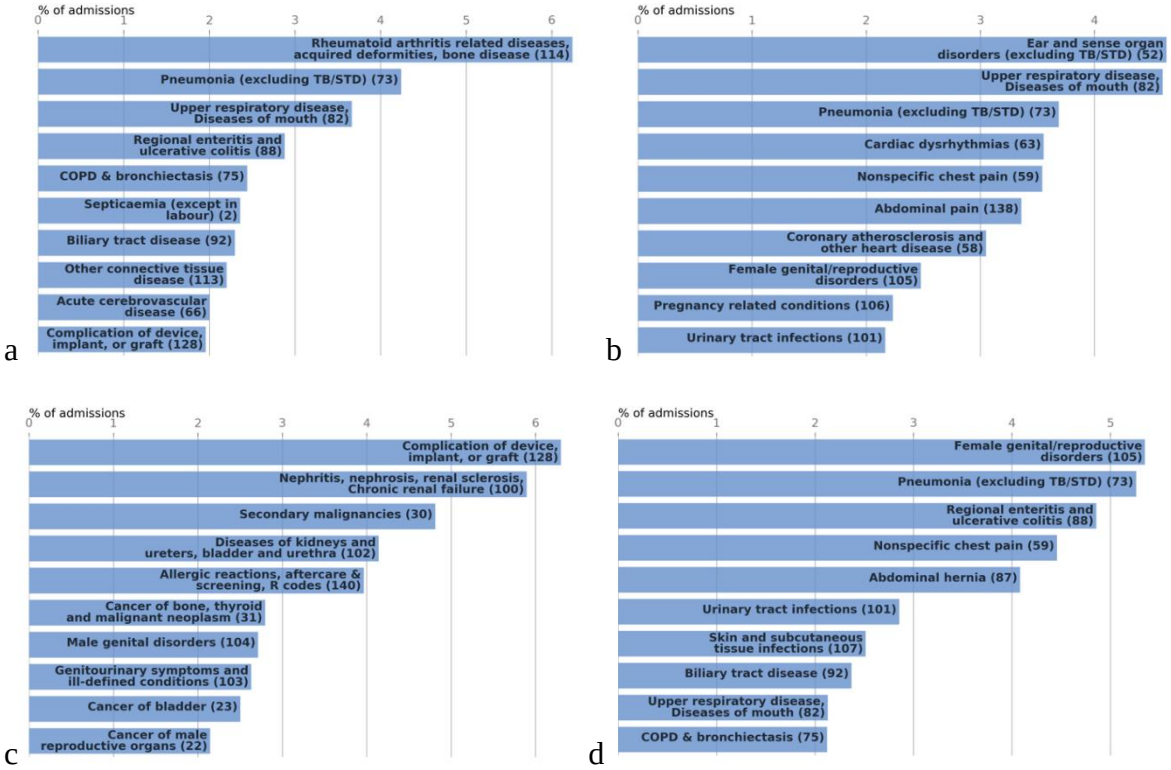


Figure SB2.

Percentage of admissions (top-10) within Summary Hospital-level Mortality Indicator (SHMI) diagnosis groups for each of the four hospitals in the test set (hospitals A, B, C, D respectively). The summary description and diagnosis group number (within brackets) are displayed for each bar. Source of SHMI group methodology: Hospital Episode Statistics (HES) data linked to Office for National Statistics (ONS) death registration data.

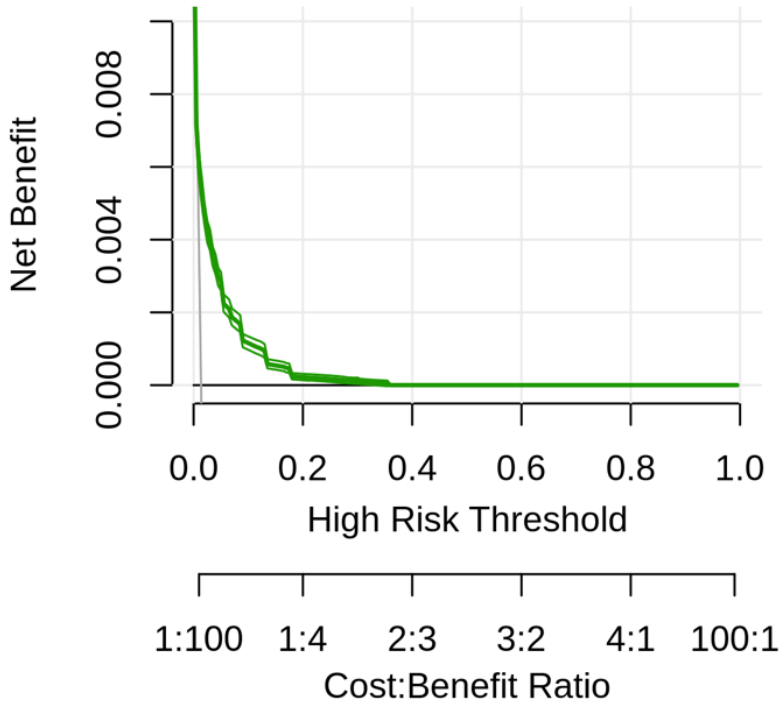


Figure SB3.

Decision curve analysis for the HAVEN model for identifying adverse events within 48 hours (green curve). The cost:benefit ratio corresponds to each threshold probability. Here, the cost can be thought of as the workload associated with reviewing patients who do not have an in-hospital cardiac arrest or unplanned ICU transfers. The curve (including 95% confidence intervals) is compared to the curves of reviewing no patients (grey line, at net benefit of 0) and all patients (light-grey line). HAVEN has a positive net benefit across low risk thresholds, performing better than reviewing none or all patients. For high risk thresholds (> 0.35), or equivalently, for high cost:benefit ratios, the net benefit of HAVEN is 0, and no different than identifying no patients at risk of deterioration. Decision curves were obtained with the *rmda* R package.

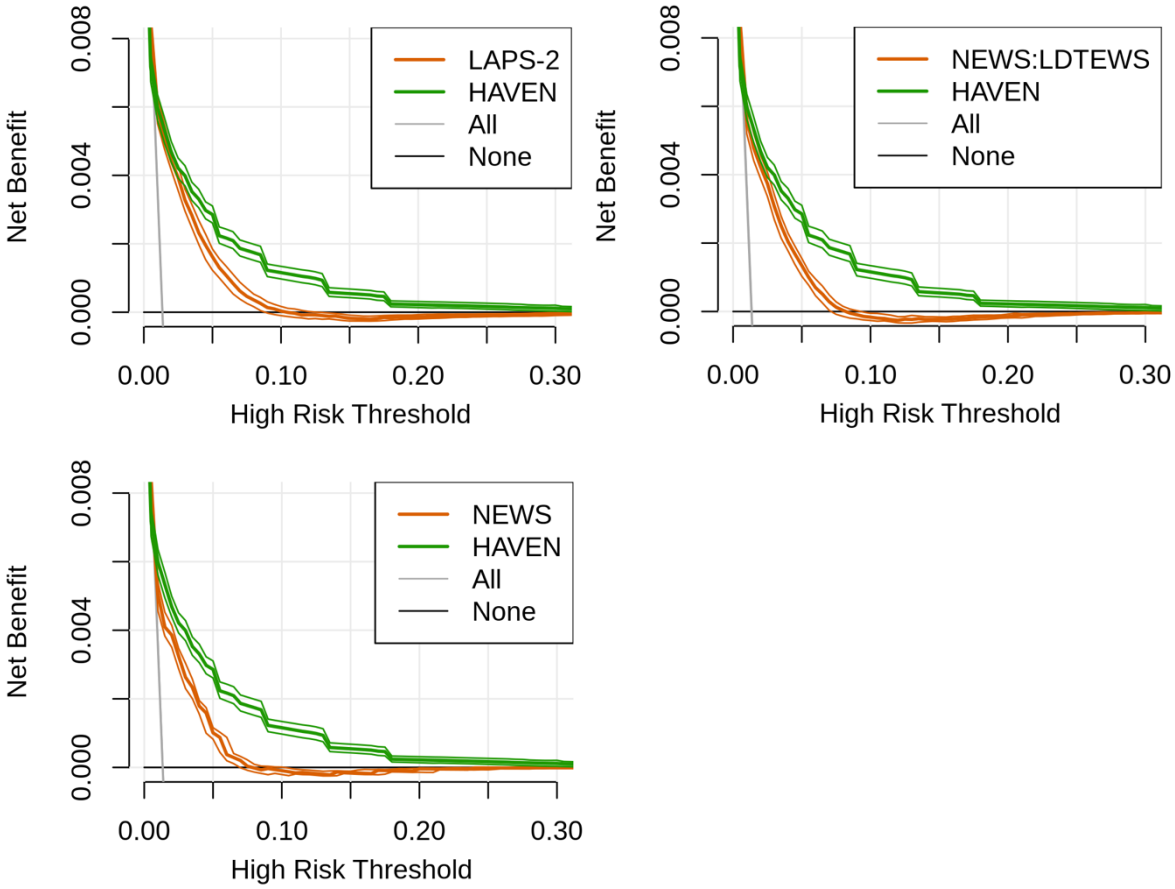


Figure SB4.

Decision curves (with 95% confidence intervals) for the default strategies (identifying /reviewing none and all patients), and the HAVEN model (green curve) compared to LAPS-2 (top-left), NEWS:LDTEWS (top-right) and NEWS (bottom) for identifying adverse events within 48 hours. As LAPS-2, NEWS:LDTEWS and NEWS do not output a risk prediction, we fitted logistic regression models to obtain the risk estimates. HAVEN has a higher net benefit than all three scoring systems for risk thresholds up to 0.3. For a cost:benefit ratio of 1:10, HAVEN has a significantly higher positive net benefit than LAPS-2 (with a net benefit of 0). For risk thresholds > 0.10, the three comparative scoring systems show a net benefit that is no better than the default strategy of reviewing no patients. Decision curves were obtained with the *rmda* R package.

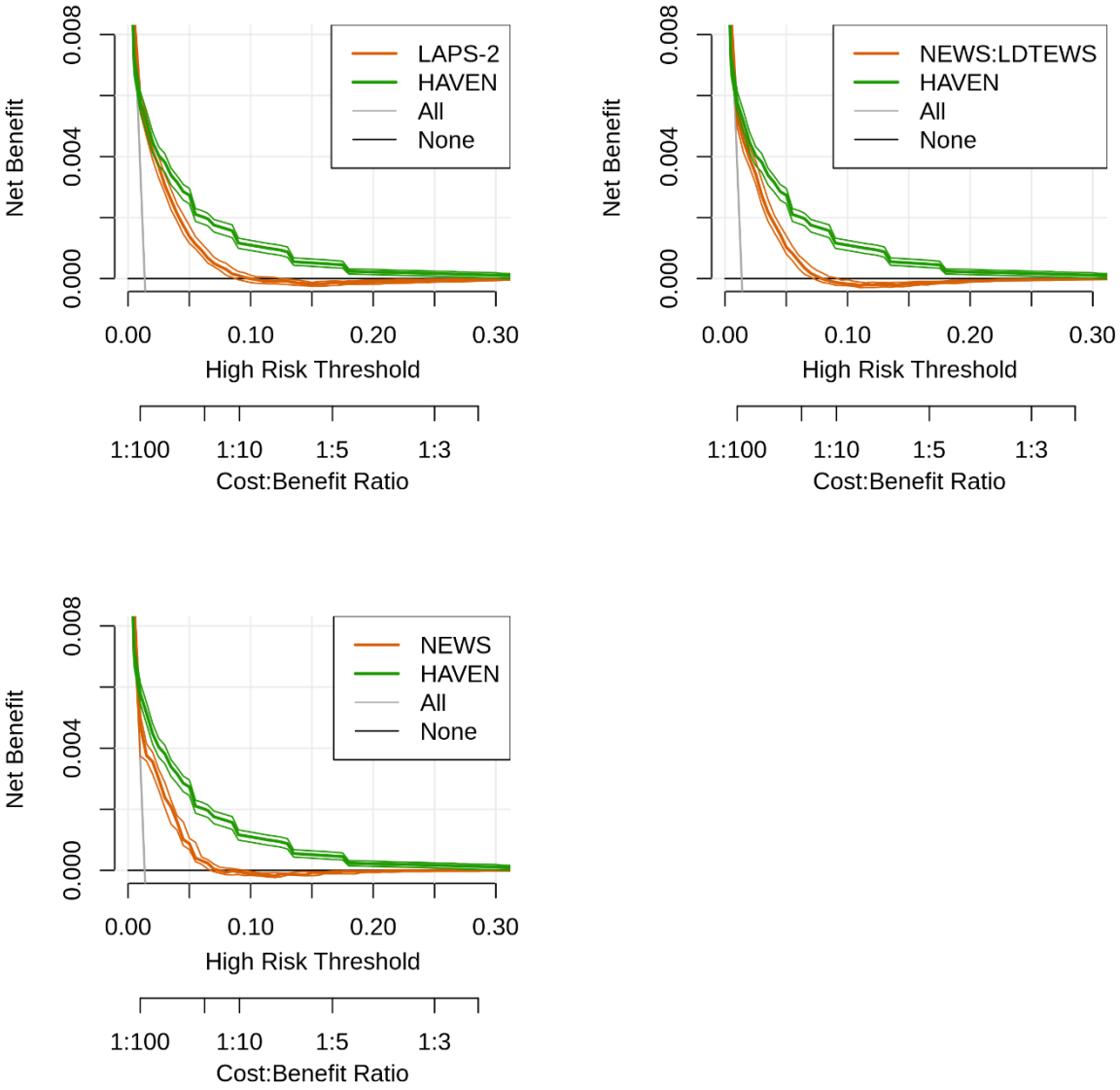


Figure SB5.

Decision curves (with 95% confidence intervals) for the default strategies (identifying /reviewing none and all patients), and the HAVEN model (green curve) compared to LAPS-2 (top-left), NEWS:LDTEWS (top-right) and NEWS (bottom) for identifying adverse events within 24 hours. As LAPS-2, NEWS:LDTEWS and NEWS do not output a risk prediction, we fitted logistic regression models to obtain the risk estimates. HAVEN has a higher net benefit than all three scoring systems for risk thresholds up to 0.3. For a cost:benefit ratio of 1:10, HAVEN has a significantly higher positive net benefit than LAPS-2 (with a net benefit of 0). For risk thresholds > 0.10, the three comparative scoring systems show a net benefit that is no better than the default strategy of reviewing no patients. Decision curves were obtained with the *rmda* R package.

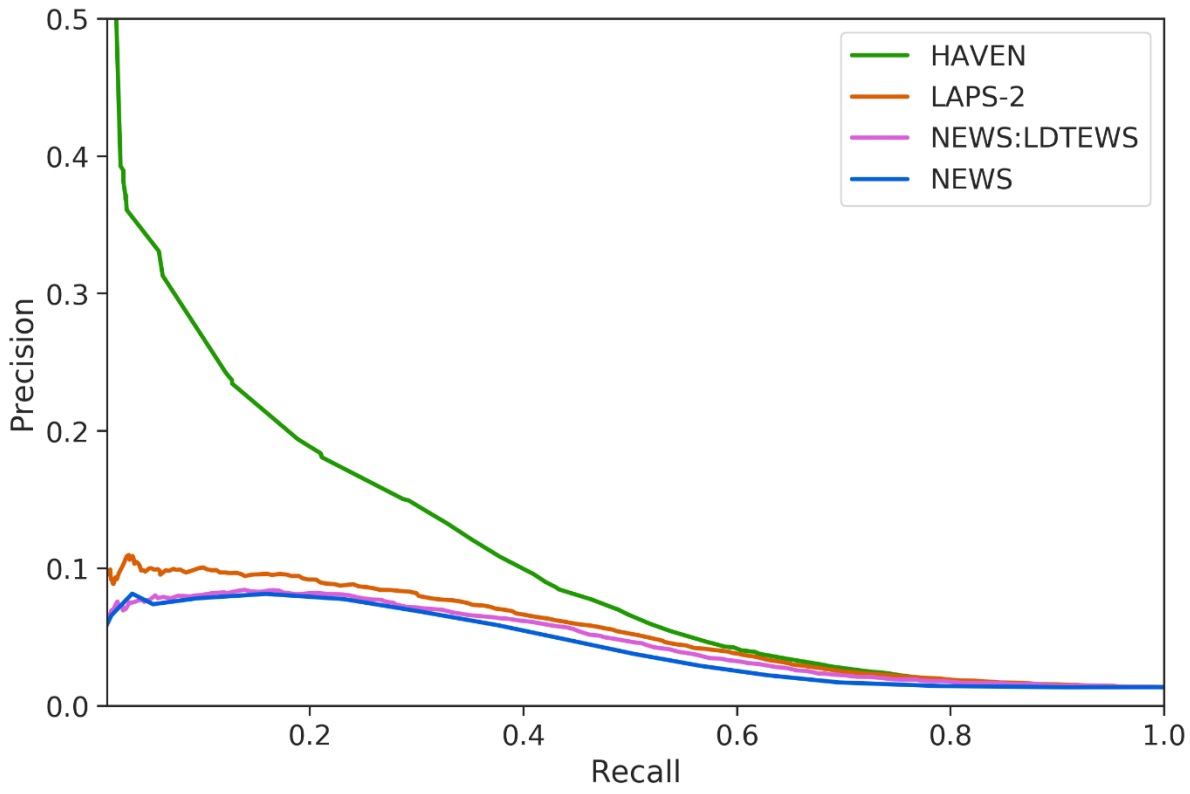


Figure SB6.

Patient admission level Precision-Recall (PR) curves for HAVEN, LAPS-2, NEWS:LDTEWS and NEWS scores, when identifying adverse events within 48 hours. As also shown in the Figure SB1, only HAVEN is able to identify events at precision levels over 0.1 (10%).

Section C – Comparative scoring systems

In this section, we describe the details of other scoring systems used in this study and how they were implemented using the information included in our database. We also detail the components and weightings of the individual scores, as they were implemented in this work.
The early warning score (EWS) systems evaluated in this study include the modified EWS (MEWS) (51), the standardised EWS (SEWS) (52), the National EWS (NEWS) (53), the centile-based EWS (CEWS) (50), and the cardiac arrest risk triage score (54). Their components and weightings are shown in Tables SC1 to SC5. The core set of variables that are used in these scoring systems comprises heart rate (measured in beats per minute), respiratory rate (in breaths per minute), systolic and diastolic blood pressure (in mmHg), body temperature (in □C), oxygen saturation (or SpO_2, in %), an assessment of neurological status (using the Alert-Verbal-Painful-Unresponsive or AVPU scale), a flag for whether the patient was receiving supplementary oxygen at the time of the SpO_2 measurement (denoted "Inspired O_2"), and the age of the patient at admission to hospital.

For all EWSs, we considered the most recent measurement in order to compute each aggregate score. For variables that were missing (or in case of an incomplete vital-sign observation set), we used the most recent value of each measurement. For the remaining missing data, the corresponding normal values were imputed, as it is the customary convention in many severity scores.

We also considered the NEWS:LDTEWS scoring system (13), which combines NEWS with a decision tree-based early warning score (LDTEWS) based on routine laboratory test results (see Table SC6). Missing data were dealt with in the same way as for the previous EWSs.

The regression-based methods considered in this study include the electronic CART (eCART) (55) and the laboratory-based acute physiology score (LAPS-2) (56) systems.

eCART includes two prediction models, one for predicting cardiac arrest, and one for predicting transfer to ICU (see table SC7). Predictor variables include vital signs (temperature, heart rate, blood pressure, SpO_2), mental status (AVPU scale), laboratory test results (white blood cell count, haemoglobin, blood urea nitrogen, anion gap, platelets count, potassium), age at admission and an indicator of whether a previous visit to the ICU occurred. We report the

results for the best performing model (i.e., the set of coefficients which yielded the best performance, as given by the ROC AUC).

LAPS-2 includes laboratory tests, vital signs, administrative and demographic information, and some interaction terms. This scoring system employs two steps for calculating the final score. In the first step, a regression model (a preliminary model), as shown in Table SC8, is used to subdivide the population into two risk groups based on their predicted mortality at admission to the hospital: a low risk group (< 6%), and a high risk group (≥ 6%). Once the two risk groups are defined, a point scoring scheme (which resulted from a second regression model) is used for determining the final aggregate score (see Table SC9). In this second score, certain variables with missing data are handled in a different way depending on the group to which the patient was assigned, i.e., for patients in the high-risk group, points are assigned for missing data for certain variables (missing and in high risk group, or M&HR); for the remaining variables, and for patients in the low-risk group, missing data are imputed with the normal values of the corresponding variables (which yields zero points), as is customary convention in many severity scores. Our implementation of this scoring system involved using both models. For the first model, admission values were set to be the first values recorded within the first 24 hours into the admission. For the second model, the most recent data from the preceding 72 hours was used for computing the score at each evaluation time point. We note that we could not include points for arterial $PaCO_2$ and PaO_2 as arterial blood gas samples are not routinely taken on ward patients.

Table SC1. Modified early warning score (see main text for units of the variables included).

Table SC2. Standardised early warning score (see main text for units of the variables included).

Table SC3. National early warning score system and its variants (see main text for units of the variables included).

Table SC4. Centile-based early warning score (see main text for units of the variables included).

Table SC5. Cardiac arrest risk triage score (see main text for units of the variables included).

Table SC6. Laboratory decision-tree early warning score.

Table SC7. The electronic Cardiac Arrest Risk Triage (eCART) model for cardiac arrest and ICU transfers.

Table SC8. The Laboratory-based Acute Physiology Score, Version 2 (LAPS-2) preliminary model coefficient estimates.

Table SC9. The Laboratory-based Acute Physiology Score, Version 2 (LAPS-2) points assignment model.

Table SC1.

Modified Early Warning Score (MEWS) (51)							
	Score						
Variable	3	2	1	0	1	2	3
Heart rate	≤ 39	$40-50$	$51-59$	$60-100$	$101-110$	$111-129$	≥ 130
Resp. rate		≤ 8		$9-14$	$15-20$	$21-29$	≥ 30
Temperature		≤ 35.0		$35.1-38.0$		$38.1-39.5$	≥ 39.6
Systolic BP	≤ 69	$70-80$	$81-100$	$101-149$	$150-169$	$170-179$	≥ 180
AVPU scale				A	V	P	U

Table SC2.

Standardised Early Warning Score (SEWS) (52)							
	Score						
Variable	3	2	1	0	1	2	3
Heart rate	≤ 29	$30-39$	$40-49$	$50-99$	$100-109$	$110-129$	≥ 130
Resp. rate	≤ 8			$9-20$	$21-30$	$31-35$	≥ 36
Temperature	≤ 33.9	$34.0-34.9$	$35.0-35.9$	$36.0-37.9$	$38.0-38.9$	≥ 39.0	
Systolic BP	≤ 69	$70-79$	$80-99$	$100-199$		≥ 200	
SpO_2	≤ 84	$85-89$	$90-92$	≥ 93			
AVPU scale				A	V	P	U

Table SC3.

National Early Warning Score (NEWS) (53)							
	Score						
Variable	3	2	1	0	1	2	3

Heart rate	≤ 40		41 – 50	51 – 90	91 – 110	111 – 130	≥ 131
Resp. rate	≤ 8		9 – 11	12 – 20		21 – 24	≥ 25
Temperature	≤ 35.0		35.1 – 36.0	36.1 – 38.0	38.1 – 39.0	≥ 39.1	
Systolic BP	≤ 90	91 – 100	101 – 110	111 – 219			≥ 220
SpO_2	≤ 91	92 – 93	94 – 95	≥ 96			
Inspired O_2		Any O_2		Air			
AVPU scale				A			V, P, U

Table SC4.

Centile-based Early Warning Score (CEWS) (50)							
	Score						
Variable	3	2	1	0	1	2	3
Heart rate	≤ 42	43 – 49	50 – 53	54 – 104	105 – 112	113 – 127	≥ 128
Resp. rate	≤ 7	8 – 10	11 – 12	13 – 21	22 – 23	24 – 28	≥ 29
Temperature	≤ 35.4		35.5 – 35.9	36.0 – 37.3	37.4 – 38.3		≥ 38.4
Systolic BP	≤ 83	84 – 90	91 – 100	101 – 157	158 – 167	168 – 184	≥ 185
SpO_2	≤ 84	85 – 90	91 – 93	≥ 94			
Inspired O_2		Any O_2		Air			
AVPU scale				A	V		P, U

Table SC5.

Cardiac Arrest Risk Triage (CART) score (54)									
	Score								
Variable	0	4	6	8	9	12	13	15	22
Heart rate	≤ 109	110 – 139					≥ 140		

Resp. rate	≤ 20			21 – 23	24 – 25	26 – 29	≥ 30
Diastolic BP	≥ 50	40 – 49	35 – 39		≤ 34		
Age	≤ 54	55 – 59		≥ 60			

Table SC6.

Laboratory Decision-Tree Early Warning Score (LDTEWS) (13)						
For females			Score			
Variable	2	1	0	1	2	3
Haemoglobin, g/dL		≤ 12.0	12.1 — 14.8	≥ 14.9		
White blood cells count, 10^9/L			≤ 12.6	12.7 — 14.8	≥14.9	
Urea (serum), mmol/L			≤ 8.4	8.5 — 13.8		≥ 13.9
Creatinine (serum), µmol/L			≤ 91	92 − 157	≥ 158	
Sodium (serum), mmol/L		≤ 134	135 − 140	≥ 141		
Potassium (serum), mmol/L		≤ 3.3	3.4 − 4.5	≥ 4.6		
Albumin (serum), g/L	≤ 28	29 − 34	≥ 35			
For males			Score			
Variable	2	1	0	1	2	3
Haemoglobin, g/dL	≤ 11.1	11.2 – 12.8	≥ 12.9			
White blood cells count, 10^9/L			≤ 9.3	9.4 − 16.6	≥ 16.7	
Urea (serum), mmol/L			≤ 9.4	9.5 − 13.7		≥ 13.8
Creatinine (serum), µmol/L			≤ 114	115 − 179	≥ 180	
Sodium (serum), mmol/L	≤ 132		133 − 140	≥ 141		

Potassium (serum), mmol/L		≤ 3.7	3.8 − 4.4	4.5 − 4.7	≥ 4.8
Albumin (serum), g/L	≤ 30	31 − 34	≥ 35		

Table SC7.

Electronic Cardiac Arrest Risk Triage (eCART) (55)		
	Coefficient	
Variable	Cardiac arrest	ICU transfer
Time*****, hours	0.00	0.00
Prior ICU stay[†], yes/no	1.37	0.12
Heart Rate, min^{-1}	0.03	0.14
Respiratory Rate, min^{-1}	0.14	0.14
Diastolic BP, mmHg	-0.02	-0.01
Temperature, □C	-0.31	-0.01
SpO_2, %	0.07	-0.05
On room air, yes/no	-0.64	-0.32
Mental status[‡], AVPU	0.43	1.16
Age at admission, years	0.03	0.02
BUN (serum)[$], mg/dL	0.01	0.01
Anion gap (serum)[§], mEq/L	0.13	0.07
Haemoglobin, g/dL	-0.17	-0.01
Platelets count[#], K/µL	-0.002	-0.001
Potassium (serum)[#], mEq/L	0.17	0.13
White blood cells count[#], K/µL	0.00	0.01

***** Time: number of hours since first ward vital sign observation

[†] Prior ICU stay: whether the patient has been admitted to the ICU or not before, during that hospital admission or care episode (yes = 1, no = 0)

[‡] Mental status: The AVPU scale was encoded as A = 0, V, P, U = 1

[$] BUN [mg/dL] = 2.8 × Urea [mmol/L]

[§] Anion gap: determined as: Sodium + Potassium - (Chloride + Bicarbonate) (serum) [all in mmol/L]

Anion gap [mEq/L] = Anion gap [mmol/L]

Note: <u>not available in our databases</u>

[#] Potassium [mmol/L] = Potassium [mEq/L]

Platelet count [K/µL] = Platelet count [10^9/L]

White blood cell count [K/µL] = White blood cell count [10^9/L]

ICU: intensive care unit; BP: blood pressure; SpO_2: oxygen saturation; BUN: blood urea nitrogen.

Table SC8.

Laboratory-based Acute Physiology Score 2 (LAPS-2) Preliminary Model (56)					
Variable	Variable Range Model coefficient				
Age at admission, years	18 − 39	40 − 64	65 − 74	75 − 84	> 84
	0.00000	-0.25234	0.25894	0.48826	0.87647
Gender	Male	Female			
	0.27430	0.00000			
Emergency room visit	Yes	No			
	1.39670	0.00000			
Sodium (serum)†, mEq/L	≤ 128	129 − 131	132 − 134	135 − 145	146 − 148
	0.11980	-0.06801	-0.30494	0.00000	-0.02560
Sodium (serum)†, mEq/L	149 − 154	> 154			
	0.42071	0.58891			
Anion gap : Bicarbonate (serum) ratio‡	≤ 199	200 − 399	400 − 599	> 599	
	-0.20038	0.00000	-0.11174	0.70227	
BUN:Creatinine ratio$	≤ 7.9	8.0 − 15.9	16.0 − 23.9	> 23.9	
	0.26988	0.00000	-0.22465	0.39858	
Intercept	—				
	-4.31678				

† Sodium [mmol/L] = Sodium [mEq/L]

‡ Anion gap: determined as: Sodium + Potassium - (Chloride + Bicarbonate) (serum) [all in mmol/L]

Anion gap [mEq/L] = Anion gap [mmol/L]

Note: <u>not available in our databases</u>

$ BUN [mg/dL] = 2.8 × Urea [mmol/L] Creatinine [mg/dL] = (1/88.42) × Creatinine [µmol/L]

BUN:Creatinine ratio determined as [mg/dL] / [mg/dL]

BUN: blood urea nitrogen; Probability p calculated as $\log(p/(1-p)) = Intercept + Coefficients \times Variables$

Table SC9.

Laboratory-based Acute Physiology Score 2 (LAPS-2) (56)					
	Variable Range				
	Points				
Variable					
Sodium (serum)†, mEq/L	≤ 128	129 − 134	135 − 145	> 145	
	14	7	0	4	
Total bilirubin (serum)‡, mg/dL	≤ 1.9	2.0 − 2.9	3.0 − 4.9	5.0 − 7.9	> 7.9
	0	11	18	25	41
BUN (serum)$, mg/dL	≤ 17	18 − 19	20 − 39	40 − 79	> 79
	0	11	12	20	25
Creatinine (serum)$, mg/dL	≤ 0.99	1.00 − 1.99	2.00 − 3.99	> 3.99	
	0	6	11	5	
BUN:Creatinine ratio$	< 25	≥ 25			
	0	10			
Albumin (serum), g/dL	≤ 1.9	2.0 − 2.4	> 2.4		
	31	15	0		
Glucose (serum)‡, mg/dL	≤ 39	40 − 59	60 − 199	> 199	
	10	10	0	3	
Haematocrit, %	≤ 19.9	20.0 − 39.9	40.0 − 49.9	> 49.9	
	7	8	0	3	
White blood cells count§, K/μL	≤ 4.9	5.0 − 12.9	> 12.9		M&HR
	8	0	11		32
Arterial $PaCO_2$, mmHg	≤ 34	35 − 44	45 − 54	55 − 64	> 64
	7	0	11	13	12
Arterial PaO_2, mmHg	≤ 49	50 − 119	> 119		
	8	0	12		
Troponin I, pg/mL	< 0.01	0.01 − 0.19	0.20 − 0.99	1.00 − 2.99	> 2.99
	0	8	17	19	25
Troponin I, pg/mL					M&HR
					9
Temperature, □F	≤ 95.9	96.0 − 100.4	> 100.4		
	20	0	3		

Heart rate, min^{-1}	≤ 59	60 − 109	110 − 139	> 139	
	7	0	7	10	
Respiratory rate, min^{-1}	≤ 19	20 − 29	> 29		
	0	11	21		
Systolic BP, mmHg	≤ 74	75 − 89	90 − 119	120 − 139	140 − 159
	22	13	5	0	8
Systolic BP, mmHg	> 159				
	14				
Shock index[#]	≤ 0.64	0.65 − 0.84	> 0.84		
	0	8	17		
SpO_2, %	≤ 89	90 − 93	> 93		
	22	12	0		

[†] Sodium [mmol/L] = Sodium [mEq/L]

[‡] Bilirubin [mg/dL] = (1/17.1) × Bilirubin [µmol/L]

Glucose [mg/dL] = 18.0182 × Glucose [mmol/L]

[$] BUN [mg/dL] = 2.8 × Urea [mmol/L] Creatinine [mg/dL] = (1/88.42) × Creatinine [µmol/L]

BUN:Creatinine ratio determined as [mg/dL] / [mg/dL]

[§] White blood cell count [K/µL] = White blood cell count [109 /L]

[#] Shock index calculated as the ratio of Heart rate to Systolic BP, measured in [min^{-1} /mmHg]

M&HR: missing and in high risk group; BUN: blood urea nitrogen; BP: blood pressure; SpO_2: oxygen saturation.

Section D – HAVEN model predictors

We included the following *dynamic* physiological variables:

- Heart rate (HR), measured in beats per minute
- Respiratory rate (RR), in breaths per minute
- Systolic (SBP) and diastolic (DBP) blood pressure, in mmHg
- Temperature, in degrees Celsius
- Peripheral oxygen saturation (SpO_2), in %
- Estimated fraction of inspired oxygen ($eFiO_2$) (66)
- Level of consciousness, using the AVPU scale[1]

We also included the following *dynamic* laboratory variables:

- Albumin, in g/L
- Bilirubin (total), in □mol/L
- Alanine aminotransferase level (ALT), in U/L
- Alkaline phosphatase level (ALP), in U/L
- Calcium (adjusted), in mmol/L
- Creatinine, in µmol/L
- C-reactive protein, in mg/L
- Estimated glomerular filtration rate, in ml/min/1.73m^2
- Haematocrit, in %
- Haemoglobin, in g/dL
- Lymphocyte count, in 10^9 cells/L
- Mean corpuscular haemoglobin (MCH), in pg/cell
- Mean corpuscular haemoglobin concentration (MCHC), in g/dL
- Mean corpuscular volume (MCV), in fL
- Platelet count, in 10^9 cells/L
- Potassium, in mmol/L
- Red blood cell count (RBC)
- Sodium, in mmol/L
- Urea, in mmol/L

[1] AVPU scale corresponds to the Alert-Verbal-Pain-Unresponsive scale. Where neurological status had been assessed using the Glasgow Coma Scale (GCS), we converted the GCS value to the AVPU scale as previously described (59–61).

- Troponin I, in pg/mL
- White blood cell count (WBC), in 10^9 cells/L

We also used the following previously-defined interaction terms:

- The shock index (defined as the ratio between HR and SBP) in beats per minute per mmHg (67, 68),
- the SpO_2-to-$eFiO_2$ ratio as % (69), and
- the Urea-to-Creatinine ratio in mmol/µmol (43, 57).

Section E – Missing data

The HAVEN GBM was trained on 10,746 positive time points (within 24 hours of the primary outcome) and 2,546,053 negative time points. Therefore, the training data set was highly class-imbalanced with 237 negative observations for every positive.

Table SE1 shows the extent of missing data in the training and test sets. For each predictor used in the model, we report the number and percentage of admissions where the predictor was never measured.

Table SE1.

Variable	Training set		Test set	
	N	*%*	*N*	*%*
Demographics:				
Age	0	0.0%	0	0.0%
Gender	0	0.0%	0	0.0%
Physiological variables:				
Heart rate	207	0.1%	293	0.1%
Respiratory rate	346	0.2%	1,624	0.6%
Systolic and diastolic blood pressure	253	0.1%	346	0.1%
Temperature	506	0.2%	5,113	1.9%
Peripheral oxygen saturation	438	0.2%	399	0.2%
Estimated fraction of inspired oxygen	714	0.3%	2,663	1.0%
Level of consciousness	876	0.4%	10,918	4.1%
Laboratory variables:				
Albumin	75,783	32.9 %	83,910	31.5 %
Bilirubin (total)	81,406	35.3 %	85,800	32.2 %
Alanine aminotransferase level	77,672	33.7 %	86,892	32.6 %
Alkaline phosphatase level	75,853	32.9 %	84,602	31.8 %
Calcium (adjusted)	92,926	40.3 %	151,548	56.9 %
Creatinine	61,567	26.7 %	65,695	24.7 %
C-reactive protein	76,221	33.1 %	98,529	37.0 %
Estimated glomerular filtration rate	61,567	26.7 %	65,699	37.1 %
Haematocrit	60,829	26.4 %	64,683	24.3 %
Haemoglobin	56,982	24.7 %	64,683	24.3 %
Lymphocyte count	57,903	25.1 %	64,762	24.3 %
Mean corpuscular haemoglobin	57,373	24.9 %	64,683	24.3 %
Mean corpuscular haemoglobin conc.	57,373	24.9 %	150,590	56.6 %
Mean corpuscular volume	57,373	24.9 %	64,683	24.3 %
Platelet count	57,372	24.9 %	64,656	24.3 %
Potassium	61,544	26.7 %	65,748	24.7 %

Red blood cell count	57,374	24.9 %	64,556	24.7 %
Sodium	61,544	26.7 %	65,722	24.7 %
Urea	61,382	26.6 %	65,748	24.7 %
Troponin I	198,157	86.0 %	231,463	86.9 %
White blood cell count	57,373	24.9 %	64,656	24.3 %
Previous admissions:				
ICD-10 diagnosis codes	85,968	37.3 %	97,997	36.8 %
Other administrative information	85,853	37.3 %	97,837	36.7 %

Section F – Feature importance

We used the SHapley Additive exPlanations (SHAP) algorithm to calculate the relative "importance" of each feature in the final gradient boosting machine model (49). Shapley values were developed within cooperative game theory as a solution to the problem of fairly distributing a total payoff between a coalition of players. Lundberg and Lee described how Shapley values could be adapted to calculate feature (predictor) importance. The SHAP algorithm can be applied to any model, including those where there are complex interactions between features (49). Individual predictions are decomposed into SHAP values for each feature. The sum of these SHAP values, plus the mean prediction, equals the individual prediction. This is analogous (but not equivalent) to adding the intercept to coefficients multiplied by the feature values in a linear model.

SHAP values can be positive or negative. The overall importance of each feature is calculated by summing the absolute Shapley value over all individual predictions.

Figure SF1 shows the top 20 predictors in our HAVEN model, ranked by their importance. Figure SF2 shows a similar plot of all features, but where individual comorbidities are combined into a single value. Figure SF1 suggests that having a known "Solid tumour" or a diagnosis of "Chronic Pulmonary Disease" has more influence on the likelihood of the primary outcome, than the presence of other comorbidities. However, Figure SF2 suggests that comorbidities as a whole are the fifth most important feature.

Figure SF3 shows a “partial dependency plot” for age, calculated from individual SHAP values. Over the age of 80, age reduces the likelihood of the primary outcome. However, there is a larger variation in the SHAP values of older patients, suggesting that other predictors (e.g., comorbidities) are considered by the model for the final prediction.

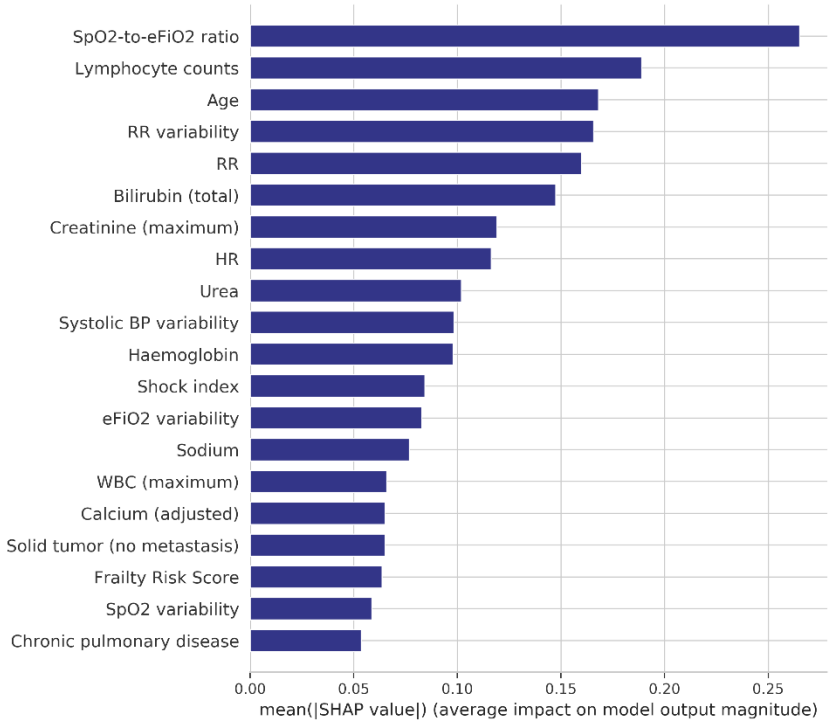


Figure SF1

Feature importance of Top 20 predictors in HAVEN GBM

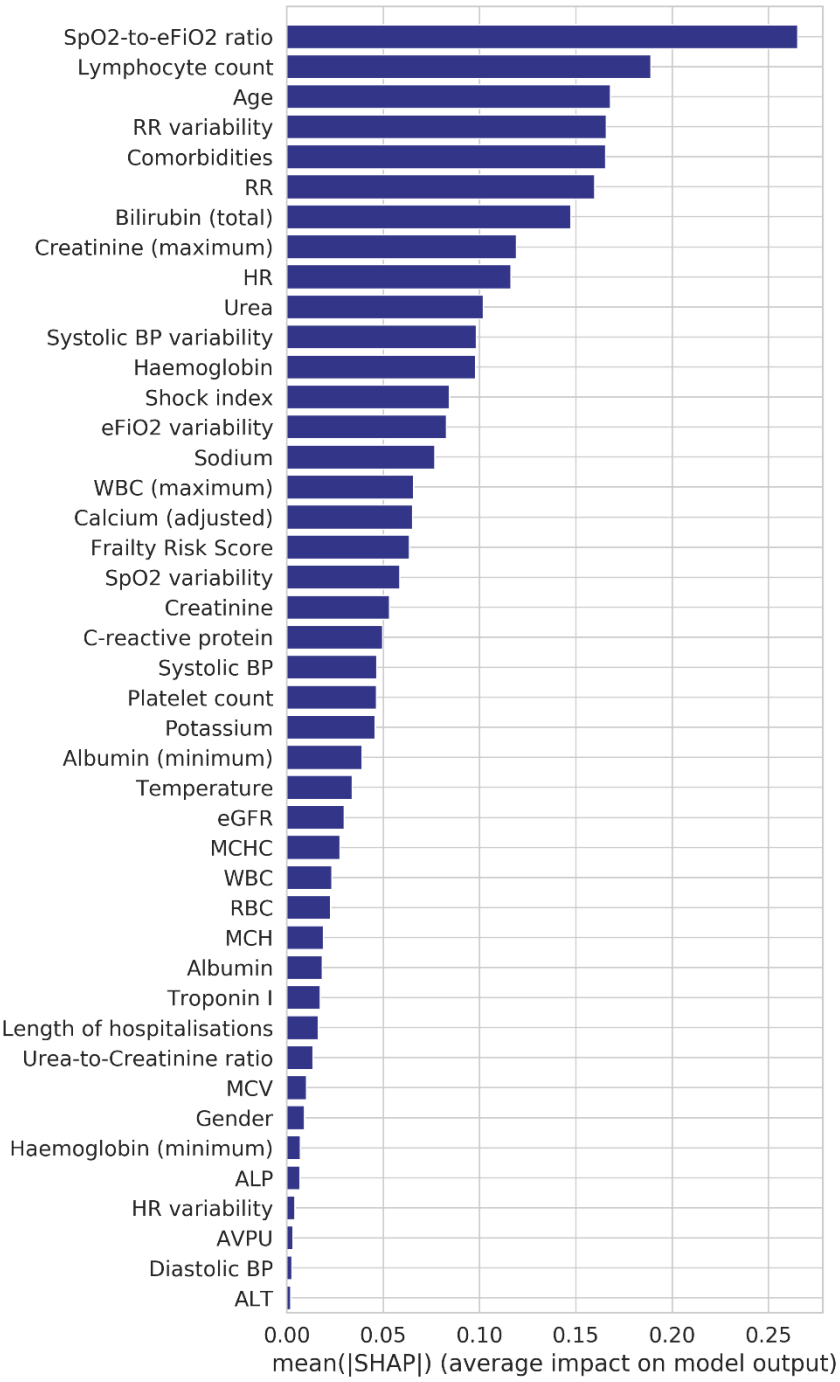


Figure SF2

Feature importance of all predictors in HAVEN GBM. Mean importance values for Elixhauser comorbidities and smoking status have been grouped into a single “comorbidities” category.

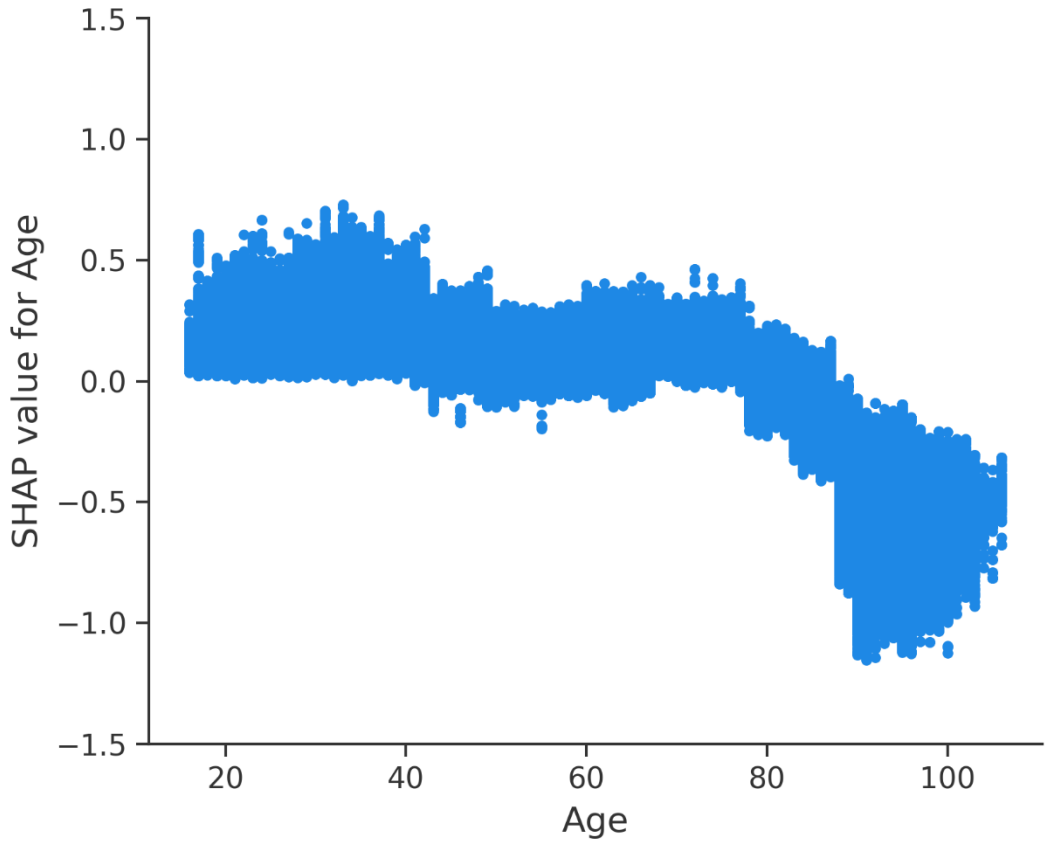


Figure SF3

Partial dependency plot of SHAP values for Age from individual predictions.

Section G – Sensitivity analysis

Table SG1 shows the performance of the HAVEN model when predicting the primary outcome in the test set at three time intervals (12, 24 and 48 hours). ROC-AUC and PR-AUC values are given for complete data and when all predictors that rely on ICD-10 coding have been set to missing. The latter simulates a scenario where no diagnostic coding is available from previous admissions. The absence of diagnostic coding results in a significant reduction in performance, suggesting that this information contributes substantially to the predictions.

	12 hours	**24 hours**	**48 hours**
ROC-AUC (Diagnostic coding set to missing)	0.916 (0.912-0.919)	0.889 (0.886-0.893)	0.866 (0.863-0.869)
ROC-AUC (Complete data)	0.921 (0.918-0.923)	0.901 (0.898-0.903)	0.881 (0.879-0.883)
PR AUC (Diagnostic coding set to missing)	0.065 (0.061-0.070)	0.073 (0.070-0.077)	0.077 (0.074-0.080)
PR AUC (Complete data)	0.074 (0.069-0.078)	0.080 (0.076-0.084)	0.082 (0.079-0.085)

Table SG1

Model performance using the test set using either complete data or where all features that rely on ICD-10 have been set to missing values. ROC AUC PR AUC values are given for predicting the risk of the primary outcome across different time windows.

References

1. Intensive Care National Audit and Research Centre (ICNARC). *Hospital length of stay pre-ICU admission and outcomes*. 2019.
2. Barwise A, Thongprayoon C, Gajic O, Jensen J, Herasevich V, Pickering BW. Delayed Rapid Response Team Activation Is Associated With Increased Hospital Mortality, Morbidity, and Length of Stay in a Tertiary Care Institution. *Crit Care Med* 2016;44:54–63.
3. National Confidential Enquiry into Patient Outcomes and Death. *Time to Intervene? A review of patients who underwent cardiopulmonary resuscitation as a result of an in-hospital cardiorespiratory arrest.* 2012.
4. Keogh B. Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report. *NHS* 2013;
5. Gerry S, Birks J, Bonnici T, Watkinson PJ, Kirtley S, Collins GS. Early warning scores for detecting deterioration in adult hospital patients: a systematic review protocol. *BMJ Open* 2017;7:e019268.
6. Smith GB, Prytherch DR, Jarvis S, Kovacs C, Meredith P, Schmidt PE, Briggs J. A Comparison of the Ability of the Physiologic Components of Medical Emergency Team Criteria and the U.K. National Early Warning Score to Discriminate Patients at Risk of a Range of Adverse Clinical Outcomes*. *Crit Care Med* 2016;44:2171–2181.
7. Pedersen NE, Rasmussen LS, Petersen JA, Gerds TA, Østergaard D, Lippert A. A critical assessment of early warning score records in 168,000 patients. *J Clin Monit Comput* 2017;32:109–116.
8. National Institute for Health and Care Excellence (NICE). *Emergency and acute medical care in over 16s: Service Delivery and Organisation*. 2018.
9. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and Early Warning Systems (EWS) for the prevention of Intensive Care admission and death of critically ill adult patients on general hospital wards. *Cochrane Database Syst Rev* 2007;CD005529.doi:10.1002/14651858.CD005529.pub2.
10. Solomon RS, Corwin GS, Barclay DC, Quddusi SF, Dannenberg MD. Effectiveness of rapid response teams on rates of in-hospital cardiopulmonary arrest and mortality: A systematic review and meta-analysis. *J Hosp Med* 2016;11:438–445.
11. McNeill G, Bryden D. Do either early warning systems or emergency response teams improve hospital patient survival? A systematic review. *Resuscitation* 2013;84:1652–

1667.

12. Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal Impact of Implemented Early Warning Score and Best Practice Alert for Patient Deterioration. *Crit Care Med* 2019;47:49–55.
13. Redfern OC, Pimentel MAF, Prytherch D, Meredith P, Clifton DA, Tarassenko L, Smith GB, Watkinson PJ. Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: Development and validation of a multivariable model. *Resuscitation* 2018;133:75–81.
14. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit Care Med* 2016;44:368–374.
15. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, Jung K, Heller K, Kale D, Saeed M, Ossorio PN, Thadaney-Israni S, Goldenberg A. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med* 2019;25:1337–1340.
16. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
17. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019;25:30–36.
18. Kang MA, Churpek MM, Zadravecz FJ, Adhikari R, Twu NM, Edelson DP. Real-Time Risk Prediction on the Wards. *Crit Care Med* 2016;44:1468–1473.
19. Dummett BA, Adams C, Scruth E, Liu V, Guo M, Escobar GJ. Incorporating an Early Detection System Into Routine Clinical Practice in Two Community Hospitals. *J Hosp Med* 2016;11:S25–S31.
20. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data. *Ann Am Thorac Soc* 2018;15:846–853.
21. Arnold J, Davis A, Fischhoff B, Yecies E, Grace J, Klobuka A, Mohan D, Hanmer J. Comparing the predictive ability of a commercial artificial intelligence early warning system with physician judgement for clinical deterioration in hospitalised general internal medicine patients: a prospective observational study. *BMJ Open* 2019;9:e032187.
22. Kipnis P, Turk BJ, Wulf DA, LaGuardia JC, Liu V, Churpek MM, Romero-Brufau S,

Escobar GJ. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform* 2016;64:10–19.

23. Sendak MP, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, Ratliff W, Balu S. A Path for Translation of Machine Learning Products into Healthcare Delivery. *EMJ Innov* 2020;doi:10.33590/emjinnov/19-00172.
24. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, Mottram A, Meyer C, Ravuri S, Protsyuk I, Connell A, Hughes CO, Karthikesalingam A, Cornebise J, Montgomery H, Rees G, Laing C, Baker CR, Peterson K, Reeves R, Hassabis D, King D, Suleyman M, Back T, Nielson C, Ledsam JR, Mohamed S. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572:116–119.
25. O'Brien C, Goldstein BA, Shen Y, Phelan M, Lambert C, Bedoya AD, Steorts RC. Development, Implementation, and Evaluation of an In-Hospital Optimized Early Warning Score for Patient Deterioration. *MDM Policy Pract* 2020;5:238146831989966.
26. Rothman MJ. The Emperor Has No Clothes. *Crit Care Med* 2019;47:129–130.
27. Paulson SS, Dummett BA, Green J, Scruth E, Reyes V, Escobar GJ. What Do We Do After the Pilot Is Done? Implementation of a Hospital Early Warning System at Scale. *Jt Comm J Qual Patient Saf* 2020;46:207–216.
28. Finlay GD, Rothman MJ, Smith RA. Measuring the modified early warning score and the Rothman Index: Advantages of utilizing the electronic medical record in an early warning system. *J Hosp Med* 2014;9:116–119.
29. Bartkowiak B, Snyder AM, Benjamin A, Schneider A, Twu NM, Churpek MM, Roggin KK, Edelson DP. Validating the Electronic Cardiac Arrest Risk Triage (eCART) Score for Risk Stratification of Surgical Inpatients in the Postoperative Setting. *Ann Surg* 2019;269:1059–1063.
30. Lindsell CJ, Stead WW, Johnson KB. Action-Informed Artificial Intelligence—Matching the Algorithm to the Problem. *JAMA* 2020;1–2.doi:10.1001/jama.2020.5035.
31. Adibi A, Sadatsafavi M, Ioannidis JPA. Validation and Utility Testing of Clinical Prediction Models. *JAMA* 2020;324:235.
32. Tarassenko L, Clifton DA, Pinsky MR, Hravnak MT, Woods JR, Watkinson PJ. Centile-based early warning scores derived from statistical distributions of vital signs. *Resuscitation* 2011;82:1013–1018.

33. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;162:W1.
34. Care Quality Commission. Oxford University Hospitals NHS Foundation Trust. at <https://www.cqc.org.uk/provider/RTH/reports>.
35. Malycha J, Bonnici T, Clifton DA, Ludbrook G, Young JD, Watkinson PJ. Patient centred variables with univariate associations with unplanned ICU admission: a systematic review. *BMC Med Inform Decis Mak* 2019;19:98.
36. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data. *Med Care* 1998;36:8–27.
37. Gilbert T, Neuburger J, Kraindler J, Keeble E, Smith P, Ariti C, Arora S, Street A, Parker S, Roberts HC, Bardsley M, Conroy S. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet* 2018;391:1775–1782.
38. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation* 2016;102:1–5.
39. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393–b2393.
40. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T, Li M, Xie J, Lin M, Geng Y, Li Y. xgboost: Extreme Gradient Boosting. 2019;
41. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. *Proc 22nd Int Conf Mach Learn - ICML '05* New York, New York, USA: ACM Press; 2005. p. 625–632.doi:10.1145/1102351.1102430.
42. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. In: Brock G, editor. *PLoS One* 2015;10:e0118432.
43. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Meltzer DO, Gibbons RD, Edelson DP. Multicenter Development and Validation of a Risk Stratification Tool for Ward Patients. *Am J Respir Crit Care Med* 2014;190:649–655.
44. Gerry S, Bonnici T, Birks J, Kirtley S, Virdee PS, Watkinson PJ, Collins GS. Early warning scores for detecting deterioration in adult hospital patients: systematic review

and critical appraisal of methodology. *BMJ* 2020;m1501.doi:10.1136/bmj.m1501.

45. Zhang Z, Rousson V, Lee W-C, Ferdynus C, Chen M, Qian X, Guo Y. Decision curve analysis: a technical note. *Ann Transl Med* 2018;6:308–308.
46. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ, Steyerberg EW. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol* 2018;74:796–804.
47. Brown M. rmda: Risk Model Decision Analysis. 2018;
48. DiCiccio TJ, Efron B. Bootstrap confidence intervals. *Stat Sci* 1996;doi:10.1214/ss/1032280214.
49. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. 2017;
50. Watkinson PJ, Pimentel MAF, Clifton DA, Tarassenko L. Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data. *Resuscitation* 2018;129:55–60.
51. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The Value of Modified Early Warning Score (MEWS) in Surgical In-Patients: A Prospective Observational Study. *Ann R Coll Surg Engl* 2006;88:571–575.
52. Paterson R, MacLeod D, Thetford D, Beattie A, Graham C, Lam S, Bell D. Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit. *Clin Med (Northfield Il)* 2006;6:281–284.
53. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. 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;84:465–470.
54. Churpek MM, Yuen TC, Park SY, Meltzer DO, Hall JB, Edelson DP. Derivation of a cardiac arrest prediction model using ward vital signs. *Crit Care Med* 2012;40:2102–2108.
55. Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP, Park S EDCMYT, Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using Electronic Health Record Data to Develop and Validate a Prediction Model for Adverse Outcomes in the Wards*. *Crit Care Med* 2014;42:841–848.
56. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P. Risk-adjusting Hospital Mortality Using a Comprehensive Electronic Record in an Integrated Health Care Delivery System. *Med Care* 2013;51:446–453.
57. Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous

measure of patient condition using the Electronic Medical Record. *J Biomed Inform* 2013;46:837–848.

58. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol* 2015;67:1142–1151.
59. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS—Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation* 2010;81:932–937.
60. Churpek MM, Yuen TC, Edelson DP. Predicting clinical deterioration in the hospital: The impact of outcome selection. *Resuscitation* 2013;84:564–568.
61. Sandroni C, D'Arrigo S, Antonelli M. Rapid response systems: are they really effective? *Crit Care* 2015;19:104.
62. Hogan H, Zipfel R, Neuburger J, Hutchings A, Darzi A, Black N. Avoidability of hospital deaths and association with hospital-wide mortality ratios: retrospective case record review and regression analysis. *BMJ* 2015;351:h3239.
63. Chukwusa E, Verne J, Polato G, Taylor R, J Higginson I, Gao W. Urban and rural differences in geographical accessibility to inpatient palliative and end-of-life (PEoLC) facilities and place of death: a national population-based study in England, UK. *Int J Health Geogr* 2019;18:8.
64. Gopalan PD, Pershad S. Decision-making in ICU – A systematic review of factors considered important by ICU clinician decision makers with regard to ICU triage decisions. *J Crit Care* 2019;50:99–110.
65. Green M, Lander H, Snyder A, Hudson P, Churpek M, Edelson D. 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;123:86–91.
66. Malycha J, Farajidavar N, Pimentel MAF, Redfern O, Clifton DA, Tarassenko L, Meredith P, Prytherch D, Ludbrook G, Young JD, Watkinson PJ. The effect of fractional inspired oxygen concentration on early warning score performance: A database analysis. *Resuscitation* 2019;139:192–199.
67. Rady MY. The role of central venous oximetry, lactic acid concentration and shock index in the evaluation of clinical shock: a review. *Resuscitation* 1992;24:55–60.
68. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, Peterson E, Tomlanovich M. Early Goal-Directed Therapy in the Treatment of Severe Sepsis and

Septic Shock. *N Engl J Med* 2001;345:1368–1377.

69. Kwack WG, Lee DS, Min H, Choi YY, Yun M, Kim Y, Lee SH, Song I, Park JS, Cho Y-J, Jo YH, Yoon H Il, Lee JH, Lee C-T, Lee YJ. Evaluation of the SpO2/FiO2 ratio as a predictor of intensive care unit transfers in respiratory ward patients for whom the rapid response system has been activated. In: Kou YR, editor. *PLoS One* 2018;13:e0201632.

Chapter 5. Towards better outcome measures for HAVEN

There are currently three main outcome measures used for the evaluation of EWS and systems like HAVEN. These outcome measures are surrogates for the deteriorated ward patient, and each has inherent limitations. Firstly, in-hospital cardiac arrests are easily measured but their frequency is rare, which limits their use for derivation and validation, even in large patient data sets. Additionally, cardiac arrests are frequently miscoded by Medical Emergency Response teams which introduces error. Secondly, in-hospital death is also easily measured but many patients progress to this outcome irreversibly, not because of missed deterioration but rather a pre-determined end of life care pathway. Finally, unplanned ICU admission is common but is influenced by clinical factors such as physiological parameters, diagnosis and prognosis. It is also influenced by logistic factors such as local ICU admission policies. Overall, models like HAVEN are subjected to the biases inherent in each during derivation, which may affect performance during clinical use. As electronic surveillance of hospital in-patients evolves there is an emerging need to define an accurate outcome measure representing clinically significant deterioration that is objective and recognisable in the digital landscape of the EPR.

The protocol outlined in this chapter is one of the multiple work programs stemming from the HAVEN project. There is a need to develop a better outcome measure. This protocol aims to describe a method to better define the deteriorated ward patient. To achieve this, we will use a literature review and validated consensus building methods. Consensus definitions will be established in three stages. Firstly, we will undertake a systematic literature review to identify existing criteria. Secondly, an international modified Delphi study will generate a ‘short list’ of candidate definitions. Finally, a Nominal Group Technique (NGT) meeting, informed by the data generated from the first two stages, will be used to complete the consensus building process.

ESTABLISHING CONSENSUS DEFINITIONS FOR THE DETERIORATED WARD PATIENT – A PROTOCOL

Authors

James Malycha MBBD FCICM[1,2,3], Chris Andersen BSc (Hon1) BA MBBS (Hon) MClinTRes FCICM[4], Oliver Redfern MBBS PhD[1], Sandra Peake BM BS BSc (Hons) FCICM PhD[2,3], Chris Subbe MBBS FRACP[5], Lukah Dykes BSc[6], Adam Phillips BSc BPharm (Hons) CHIA[6], Guy Ludbrook PhD MBBS FANZCA[2], Duncan Young BM DM FRCA FMedSci[1], Peter Watkinson MD ChB MRCP FDICM[1], Sandra Peake BM BS BSc (Hons) FCICM PhD [2,3,6]Arthas Flabouris MBBS MD FCICM FANZCA FACAsM PGDipAvMed[2], Daryl Jones BSc (Hons), MB BS, FRACP, FCICM, MD, PhD[7]

Affiliations

[1]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[2]Discipline of Acute Care Medicine, University of Adelaide, South Australia, Australia

[3]Intensive Care Unit, The Queen Elizabeth Hospital, Adelaide, South Australia, Australia

[4] Intensive Care Unit, Royal North Shore Hospital, Sydney, New South Wales, Australia

[5]Acute Medicine Department, Bangor Hospital, Bangor, Wales, UK

[6]Central Adelaide Local Health Network, Adelaide, South Australia, Australia

[7]Intensive Care Unit, Austin Hospital, Melbourne, Victoria, Australia

Keywords

Critical Care, Rapid Response System, Patient Deterioration, Outcome Measures, Consensus

Funding

PW and OR are both supported by the National Institute for Health and Research Biomedical Research Centre, Oxford. CS does consultancy work for Philips Healthcare.

Ethics

Ethics approval will be sought for the use of the patient database during the study.

Statement of Authorship

Title of paper	Establishing consensus definitions for the deteriorating ward patient – a protocol
Publication status	Submitted
Publication details	Establishing consensus definitions for the deteriorating ward patient – a protocol. J Malycha, C Andersen, O Redfern, S Peake, A Phillips, C Subbe, G Ludbrook, JD Young, P Watkinson, A Flabouris, D Jones. Submitted Aug 2021

Principal Author

Name of Principal Author	Dr James Malycha
Contribution to paper	Conceptualised the work, coordinated the authorship grou, wrote manuscript and was corresponding author
Overall percentage (%)	90%
Certification	This paper reports on original research conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Co-author contributions

Each author certifies:

- The candidate's stated contribution to the publication is accurate (as state above);
- Permission is granted for the candidate to include the publication in the thesis; and
- The sum of all co-author contributions is equal less the candidate's stated contribution

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Chris Andersen
Contribution to paper	Helped conceptualise the study, data analysis and interpretation

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Oliver Redfern
Contribution to paper	Helped conceptualise the study, data analysis, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Chris Subbe
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Mr Lukah Dykes
Contribution to paper	Data analysis, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Mr Adam Phillips
Contribution to paper	Data analysis, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Guy Ludbrook
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Duncan Young
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Sandra Peake
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Peter Watkinson
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Associate Professor Arthas Flabouris
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Daryl Jones
Contribution to paper	Evaluated and edited manuscript

Abstract

Introduction

Most patients admitted to hospital recover with treatments that can be administered on the general ward. However, a small cohort deteriorate to the extent that they require augmented organ support. In observational studies evaluating this cohort, proxy outcomes are used, including unplanned transfer from general ward to the Intensive Care Unit (ICU), cardiac arrest and death. However, these outcomes measures have limitations.

Aims

This protocol aims to describe a method to better define the deteriorated ward patient. To achieve this, we will use a literature review and validated consensus building methods.

Methods

1. We will undertake a systematic literature review to identify existing definitions.
2. An international modified Delphi study will generate a 'short list' of candidate definitions.
3. A Nominal Group Technique (NGT) meeting, informed by the data generated from the first two stages, will be used to complete the consensus building process.

Discussion

The results of the study will be made available to international researchers. It is anticipated the definitions will then be evaluated and iterated by different research teams. These results will inform the international research community on the relevance of the definitions and their potential usefulness. Ideally, the definitions will hasten the development and improve the performance of automated, Electronic Medical Record (EMR) linked, digital models that accurately predict which general ward patients will require augmented organ support (as opposed to predicting death, cardiac arrest or unplanned ICU admission).

Introduction

Most patients admitted to hospital recover with treatments that can be administered on the general ward. However, a small cohort within this population deteriorate to the extent that they require augmented organ support **(Figure 1).** (1) In observational studies evaluating this cohort, proxy outcomes are most often used. These include unplanned transfer from the general ward to the Intensive Care Unit (ICU), cardiac arrest and death. (3, 4) However, the decision to transfer patients to the ICU is dependent on multiple factors, including personalised advance care directives, clinician opinion, local care escalation protocols such as Early Warning Score (EWS) systems, and the availability of ICU resources. (5) These factors introduce subjectivity and variability when ICU admission is chosen as an outcome measure, hindering evaluation of interventions designed to improve the care of these patients. (6–11)

Cardiac arrest and death are well-defined and easily measured but are often a very late marker of deterioration. Additionally, cardiac arrest frequency is rare, which limits its use for derivation and validation processes, even in large patient data sets. We aim to define this deteriorated ward patient cohort more accurately, using the time-point of when the need for augmented organ support first occurs. We will use validated consensus methods to generate the definitions using a diverse international panel of stakeholders.

Aim and Objectives

This protocol aims to describe a method to better define the deteriorated ward patient. To achieve this, we will use a literature review and validated consensus building methods.

Methods

Consensus definitions will be established in three stages. Firstly, we will undertake a systematic literature review to identify existing definitions for clinical deterioration (12) Secondly, an international modified Delphi study will generate a ‘short list’ of candidate definitions. Finally, a Nominal Group Technique (NGT) meeting, informed by the data generated from the first two stages, will be used to complete the consensus building process. Definitions are expected to be organ system specific and will not be designed as real-time adjuncts to clinical decision

making. Both Delphi and NGT are validated methods for establishing consensus in health care settings. (13–15)

Stage 1 - Literature Review

The Preferred Reporting of Observational Studies and Meta-Analysis (PRISMA) guidelines will be used to conduct a literature review on current definitions for the deteriorating ward patient.(12) Data sources will include MEDLINE, EMBASE, CINAHL and CENTRAL (for full names see Abbreviations section). Additional papers will be included from the references of review articles. An example of the search criteria is included in the **Table 1**

Stage 2 - Delphi Study

Participants

We aim to include 60 participants in the Delphi study. Participants will be recruited through the International Society for Rapid Response Systems, the International Forum for Acute Care Trialists and relevant national societies. No formal inclusion criteria will be used however potential participants will be considered based on relevant clinical and research experience, with the aim of ensuring participants are representative of eventual end-users. These will include hospital-based clinical staff working in regional, rural and metropolitan hospitals as well as non-clinician content experts such as researchers and digital health specialists. A small number of health consumer representatives will also be recruited.

Round 1 – Establishing initial definitions (time frame: two months)

Results of the literature review and a list of potential domains, variables and/or parameters will be distributed via email to participants. Participants will provide structured feedback on the merits (or otherwise) of each item. These will then be coalesced into an initial list of potential definitions. Any missed items will be submitted to the process for consideration.

Round 2 – Ranking potential definitions (time frame: two months)

Participants will rank each potential definition using a 9-point Likert System. Consensus will be defined as 70% of respondents classifying the definitions as ‘critical’ (score of 7 - 9) and less than 15% determining the definition to be ‘not relevant’ (score 1 - 3). The results will be aggregated. Any criteria achieving a score of > 70% ‘not relevant’ will be removed.

R*ound 3 – Refining aggregated results (time frame: two months)*

Aggregated results will be presented to each participant. Definitions that remain, but that have not yet achieved consensus, will be rescored. These results will then be aggregated, and the list finalised. Any definitions that have not achieved consensus after three rounds of scoring will be excluded.

Round 4 – Generating thresholds (time frame: two months)

Participants will propose one threshold for each organ specific definition with an evidence-based justification for the threshold.

Stage 3 - Nominal Group Technique/Consensus meeting (time frame: 1 day)

Nominal Group Technique (NGT) is a validated method for establishing consensus on a specific issue or range of related issues. (14) The NGT meeting will aim to include a diverse range of clinical stakeholders. The target number of participants will be 15 - 20. Participants will be selected and invited using the same process as described for the Delphi. Participants need not have been involved in the first two stages of the study to take part.

A trained facilitator will lead NGT participants through the structured multi-stage process: Firstly, participants will be presented with an overview of the NGT meeting rationale and aim. Next, participants will be presented with the results of the systematic review and the Delphi process. Participants will then spend 10 - 15 minutes writing a list of bullet point reflections and opinions on the definitions provided, including an opportunity to advocate for additional relevant data not previously included. The facilitator will then get participants to list one reflection/opinion that is yet to be presented. Each original point will be transcribed onto a screen or whiteboard, so all participants can consider and review. This may take several rounds until opinions are exhausted (the aim being to enable all participants to express their views and prevent specific participants having a disproportionate influence).

Participants will then place each definition into two columns: one for inclusion and one for exclusion. The results of this activity will be tabulated and presented. Consensus will be confirmed if more than 70% of participants support its inclusion or exclusion. (16) If there is a lack of consensus on a definition, then the contentious item will be taken back to the group

for reappraisal and repeat voting until either consensus or stalemate (two additional voting rounds without consensus) is reached.

The final stage of the NGT will determine the thresholds (if required) for each of the definitions. Participants will write down opinions/reflections on potential thresholds and these will be collated with each original perspective transcribed. Participants will then provide specific thresholds for relevant definitions; these results will be tabulated, and discussion will be encouraged. The facilitators will present numerous potential thresholds based on the feedback and these will again be voted on. The final set of definition thresholds will be presented to the group and pending agreement, recorded for subsequent publication.

Dissemination

Results generated from this study will be disseminated through publication and presentation at national and international scientific meetings.

Discussion

In this protocol, we have described a method using international expert consensus to define the deteriorated ward patient as the time-point that the need for augmented organ support first occurs. To our knowledge this is the first study to undertake this research task. This research represents an important step in improving the precision of outcome measures used for the development and evaluation of future clinical deterioration prediction models. The proposed work is challenging. It aims to use consensus building methods that are current best practice. The development of the definitions will be an iterative process.

Once published, the results of the study will be made available to international researchers. It is anticipated the definitions will then be evaluated and iterated in observational studies by different research teams. These results will inform the international research community on the relevance of the definitions and their potential usefulness. Ideally, the definitions will hasten the development and improve the performance of automated, EMR linked, digital models that accurately predict which general ward patients will require augmented organ support (as opposed to predicting death, cardiac arrest or unplanned ICU admission).

Declarations

Acknowledgments

JM would like the acknowledge the University of Adelaide, Department of Acute Medicine, who are administering his PhD.

Contributors

JM and CA designed the study. JM, CA and DJ undertook the methodological planning. JM and CA wrote the manuscript. All authors commented on successive drafts and approved the final manuscript.

Funding

PW and OR are both supported by the National Institute for Health and Research Biomedical Research Centre, Oxford. CS does consultancy work for Philips Healthcare.

Competing interests

None

Patient consent for publication

N/A

Ethics approval

Ethics approval will be sought for the use of the EMR patient database during the study.

Provenance and peer review

Open access

Abbreviations

MEDLINE – Medical Literature Analysis and Retrieval System Online

CINAHL – Cumulative Index to Nursing and Allied Health Literature

EMBASE – Excerpta Medica database

CENTRAL – Cochrane Index to Nursing and Allied Health Literature and the Cochrane Central Register of controlled trials

References

1. Orosz J, Bailey M, Udy A, et al.: Unplanned ICU Admission from Hospital Wards after Rapid Response Team Review in Australia and New Zealand. *Crit Care Med* 2020; 48:E550–E556
2. Blanch L, Abillama FF, Amin P, et al.: Triage decisions for ICU admission: Report from the Task Force of the World Federation of Societies of Intensive and Critical Care Medicine. *J Crit Care* 2016; 36:301–305
3. Smith MEB, Chiovaro JC, O'Neil M, et al.: Early Warning System Scores for Clinical Deterioration in Hospitalized Patients: A Systematic Review. *Ann Am Thorac Soc* 2014; 11:1454–1465
4. Alam N, Hobbelink EL, van Tienhoven AJ, et al.: The impact of the use of the Early Warning Score (EWS) on patient outcomes: A systematic review. *Resuscitation* 2014; 85:587–594
5. Nunnally M, Fowler CS, Bailey H, et al.: ICU Admission, Discharge, and Triage Guidelines. 2016.
6. Churpek MM, Yuen TC, Edelson DP: Predicting clinical deterioration in the hospital: The impact of outcome selection. *Resuscitation* 2013; 84:564–568
7. Churpek MM, Carey KA, Dela Merced N, et al.: Validation of Early Warning Scores at Two Long-Term Acute Care Hospitals. *Crit Care Med* 2019; 47:e962–e965
8. Kipnis P, Turk BJ, Wulf DA, 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; 64:10–19
9. O'Brien C, Goldstein BA, Shen Y, et al.: Development, Implementation, and Evaluation of an In-Hospital Optimized Early Warning Score for Patient Deterioration. *MDM Policy Pract* 2020; 5:1–10
10. Escobar GJ, Laguardia JC, Turk BJ, et al.: Early detection of impending physiologic deterioration among patients who are not in intensive care: Development of predictive models using data from an automated electronic medical record. *J Hosp Med* 2012; 7:388–395
11. Alvarez CA, Clark CA, Zhang S, et al.: Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. *BMC Med Inform Decis Mak* 2013; 13:28
12. Moher D, Liberati A, Tetzlaff J, et al.: Systematic Reviews and Meta-Analyses: The

PRISMA Statement. *Annu Intern Med* 2009; 151:264–269

13. Sinha IP, Smyth RL, Williamson PR: Using the Delphi Technique to Determine Which Outcomes to Measure in Clinical Trials: Recommendations for the Future Based on a Systematic Review of Existing Studies. *PLoS Med* 2011; 8:e1000393
14. Grech C: Use of consensus development to establish national research priorities in critical care. *Aust Crit Care* 2000; 13:163
15. Jones J, Hunter D: Qualitative Research: Consensus methods for medical and health services research. *Bmj* 1995; 311:376
16. Rankin NM, McGregor D, Butow PN, et al.: Adapting the nominal group technique for priority setting of evidence-practice gaps in implementation science. *BMC Med Res Methodol* 2016; 16:1–9

Tables and Figures

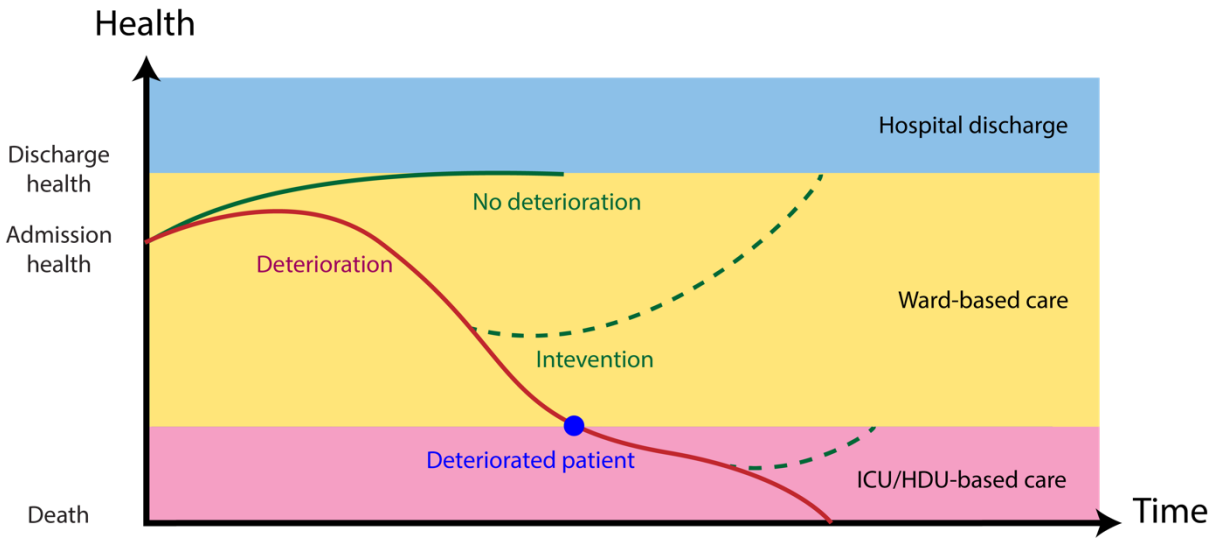


Figure 1.

A schematic representation of potential trajectories for hospitalised patient. Most patients are expected to progress along the green line. However, in a small cohort significant deterioration will occur. This may be subject to early intervention or will reach an end point at which they are no longer suitable for management in a ward environment and will be determined to have 'deteriorated'.

Table 1.

Step	Term	Studies
1	INTENSIVE CARE UNITS/	49151
2	CRITICAL CARE/st	3674
3	("intensive care" or ICU* or ITU*).ti.	46127
4	1 or 2 or 3	74307
5	PATIENT ADMISSION/st 891	891
6	TRIAGE/st 1205	1205
7	PATIENT SELECTION/ 60959	60959
8	(admission* or admit* or access* or triage*).ti. 96353	96353
9	5 or 6 or 7 or 8 157554	157554
10	4 and 9 4318	4318
11	(Criteria or assessment or optim* or survey or decision* or evaluat* or consensus or standard* or measure* or algorithm* or tool* or instrument* or guideline* or framework* or method* or strateg*).ti.	2114207
12	10 and 11	555

Table 1. Example search criteria using MEDLINE

Chapter 6. Automated quantification of the clinical workload associated with systems like HAVEN

Research plans are progressing to implement HAVEN into the South Australian ward environment via feasibility and implementation studies. Part of this research will involve an evaluation of health economic and operational performance of HAVEN (note clinical outcome measures will also be evaluated). Currently there are no published methods describing how to reliably quantify baseline work for clinicians managing deteriorating ward patients in an automated way. As a first step towards understanding how this might be achieved, a study was performed to evaluate whether Real-time Location Devices (RTLDs) can reliably quantify clinical staff workload. RTLDs are portable devices which communicate with Wi-Fi network access points to determine position. This study tracked ICU registrar 'off unit' activity over twelve months. Overall, the system demonstrated it could reliably and automatically track and record ICU registrar activity. This provided accurate and contemporary data on clinical demand coming from outside the ICU. This method will form part of future health economic and operational analysis of implementation of the HAVEN system.

USING REAL-TIME LOCATION DEVICES (RTLD) TO QUANTIFY OFF-UNIT ADULT INTENSIVE CARE REGISTRAR WORKLOAD: A ONE-YEAR TERTIARY NHS HOSPITAL PROSPECTIVE OBSERVATIONAL STUDY

Authors

James Malycha MBBD FCICM[1,2,3], Daniel Murphy BMBCh BA[4], Graham Barker MBBS FRCA FFICM[5], Guy Ludbrook PhD MBBS FANZCA[2], Duncan Young BM DM FRCA FMedSci[1], Peter Watkinson MD ChB MRCP FDICM[1,5]

Affiliations

[1]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[2]Discipline of Acute Care Medicine, University of Adelaide, South Australia, Australia
[3]Intensive Care Unit, The Queen Elizabeth Hospital, Adelaide, South Australia, Australia
[4]Medical School, University of Oxford, Oxford, UK
[5]Intensivc Care Unit, The John Radcliffe Hospital, Oxford, UK

Keywords

Intensive care unit, critical care, medical workload, real time location device

Funding

No funding was required for this study

Ethics

Health Research Authority (South Central – Oxford C, Research Ethics Committee (REC), southcentral-oxfordc@nhs.net) approval was sought but not required as the study was deemed a quality improvement project. All documents and data were stored securely and only accessible by study staff and authorised personnel. The study complied with the Data Protection Act

Statement of Authorship

Title of paper	Using real-time location devices (RTLD) to quantify off-unit adult intensive care registrar workload: a 1-year tertiary NHS hospital prospective observational study
Publication status	Published
Publication details	Malycha J, Murphy D, Barker G, Ludbook G, JD Young, P Watkinson. Using real-time location devices (RTLD) to quantify off-unit adult intensive care registrar workload: a 1-year tertiary NHS hospital prospective observational study. J Clin Monit Comput 2019; 3–7

Principal Author

Name of Principal Author	Dr James Malycha
Contribution to paper	Conceptualised the work, collected and analysed the data, wrote manuscript and was corresponding author
Overall percentage (%)	90%
Certification	This paper reports on original research conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Co-author contributions

Each author certifies:

- The candidate's stated contribution to the publication is accurate (as state above);
- Permission is granted for the candidate to include the publication in the thesis; and
- The sum of all co-author contributions is equal less the candidate's stated contribution

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Daniel Murphy
Contribution to paper	Data analysis and interpretation

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Dr Graham Barker
Contribution to paper	Helped conceptualise the study, evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Guy Ludbrook
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Duncan Young
Contribution to paper	Evaluated and edited manuscript

Verified by:	Principal Supervisor, Professor Guy Ludbrook, Adelaide School of Medicine, The University of Adelaide
31/08/2021	
Name of Co-Author	Professor Peter Watkinson
Contribution to paper	Evaluated and edited manuscript

Abstract

Introduction

UK national guidelines state deteriorating or at risk hospital ward patients should receive care from trained critical care outreach personnel. In most tertiary hospitals this involves a team led by an Intensive Care Unit (ICU) registrar. The ICU registrar must also review patients referred for possible ICU admission. These two responsibilities require work away from the ICU. To our knowledge the burden of this work has not been described, despite its importance in ICU workforce management and patient safety.

Methods

A 12-month, prospective, observational study was carried out. The primary outcome measure was ICU registrar time spent on and off-unit. The study participants were senior and junior registrars on the rota of the 16 bed, Adult Intensive Care Unit at the John Radcliffe Hospital in Oxford. To measure their work patterns, this study used AeroScout 'T2' Real Time Location Device (RTLD) tags (Stanley Healthcare, Swindon).

Results

In our hospital, senior and junior ICU registrars spend roughly one fifth of their time off-unit, half of which is spent in ED. This workload combines to leave the unit unattended at night up to 10% of the time.

Conclusion

RTLDs provide a reliable, automated method for quantifying ICU registrar off-unit work patterns. This method may be adopted for quantifying other clinical staff work patterns in suitably equipped hospital environments.

Introduction

Current guidelines stipulate that National Health Service (NHS) hospitals must have appropriate Intensive Care Unit (ICU) staffing to ensure safe on-unit and off-unit patient care. (1) The majority of ICU registrar work is generated on-unit through admitting, managing and discharging critically ill patients (a registrar is a junior doctor who has completed foundational training, usually two years, and is in training in a specialty area of medicine or surgery). However, the United Kingdom (UK) National Outreach Forum Operational Standards and Competencies for Critical Care Outreach Services state deteriorating hospital ward patients should receive care from trained critical care outreach personnel. In most NHS hospitals this involves a team led by an ICU registrar. (2) The ICU registrar must also review ward patients referred for possible ICU admission. (3) These two responsibilities require work away from the ICU. To our knowledge this work has not been quantified in the literature, despite its importance in ICU workforce management and patient safety. This study used an automated method to evaluate the work locations of ICU registrars within a tertiary NHS hospital over a 12-month period.

Methods

The 12-month, prospective, observational study was carried out in accordance with the Revised Standards for Quality Improvement Reporting Excellence (SQUIRE 2.0) guidelines. (4) The study ran from April 1st 2017 to March 31st 2018.

Intervention

Real-time Location Devices (RTLDs) are small portable devices that can communicate with wireless (wifi) data networks. This study used AeroScout ‘T2’ RTLDs, (Stanley Healthcare, Swindon). The RTLDs determine their position by triangulating the signals from wireless access points (WAPs) and pass this information to the AeroScout software. The location of each WAP is premapped into the AeroScout software, so the RTLD position within the hospital building can be determined. The RTLD location is updated every five minutes. RTLDs are usually used to determine the location of mobile assets (such as portable physiological monitors). In this study the RTLDs were attached to the “baton” pagers carried by the registrars which are handed on at the end of each shift.

Measures

The primary outcome measure was ICU registrar time spent on and off the ICU. The study participants were registrars on the senior and junior tiers of the rota for the 16 bed, Adult Intensive Care Unit of the John Radcliffe Hospital, which is part of the Oxford University Hospitals National Health Service Foundation Trust (OUHNHSFT).

Analysis

Data analysis was carried out using Python version 2.7 (Python Software Foundation, Wilmington, USA).

Ethical considerations

Informed consent was not required, but all registrars were made aware of the study. Data were collected automatically in real time and securely stored within the hospital's Aeroscout system. Access to the data was granted to JM only.

Validation

The tags were tested during the study by placing them in predesignated locations and times and cross-referencing these with the location data.

Results

Figure 1 shows the results of binary (i.e. on-unit/off-unit) analysis carried out on ICU doctor location during each of the shifts per 24 hours. During the day shift (0830 – 1800) the adult ICU was staffed by both the senior and junior registrars as well as between two and five additional doctors of varying seniority. During the evening shift (1800 – 2100) and the night shift (2100 – 0830) the ICU was staffed by only the senior and junior registrars. In the latter two shifts, the ICU was therefore left unattended if both registrars were called off-unit

Off-unit location analysis showed that during the day shift (0800 - 1830), on average, the senior and junior registrars spent 7% of their time in the Emergency Department (ED)/Emergency Assessment Unit (EAU) and 8% in the general wards. During the evening shift (1830 – 2100), on average, both registrars spent 10% in the ED/EAU and 7% in the general wards. During the

night shift (2100 – 0830), on average, both registrars spent 8% in the ED/AEU and 11% in the general wards. Variation in registrar off-unit work load between days of the week was minimal. Results are included in the Appendix (Figure 2). A detailed description of the locations making up the category 'general wards' are included in the Appendix (Figure 3a and 3b).

Discussion

Summary

The provision of intensive care in NHS hospitals increasingly involves off-unit registrar work, which is important but time consuming and expensive. We are the first to use RTLDs to quantify this activity. In our hospital, senior and junior ICU registrars spend roughly one fifth of their time off-unit, half of which is spent in the ED. This workload combines to leave the unit unattended at night up to 10% of the time. This method and these data will help inform decisions about ICU off-unit workload and staffing and may in turn improve patient safety.

Interpretation

This study established the feasibility of this method to establish staff working locations (both ICU and non-ICU) which could be applied to other hospitals. RTLD tags are already widely used in the NHS to monitor the location of portable medical devices so adapting this method may prove cost effective and efficient in those cases. Generally, a detailed understanding of when, where and how long ICU registrars spend off-unit will assist in customising staffing in ICU and in areas where ICU expertise are required (e.g., the ED). Additionally, being aware of when and how often an ICU is without a registrar with airway expertise (as was the case for up to 10% of the night shift) is of clinical importance. Locally, this data informed staffing decisions.

Limitations

The temporal resolution (5 minutes) of this study was preset by the tracking devices and could not be modified. 'Floor hopping', where the RTLD communicates with a WAP on the floor above or below is dependent on WAP layout and occurred roughly 1% of the time. This system requires Aeroscout or similar hardware and software to be installed within the hospital and collaboration with skilled hospital Information Technology technicians. These data give ICU registrar work locations and times but not work type. We made the assumption that ICU

registrars were off-unit in response to work demand alone and this may have not always been the case. Likewise, we acknowledge individual doctors will have approached the same workload in different ways. The data was anonymised so individual doctor off-unit workload patterns were not evaluated and this may have introduced bias into the results.

Conclusions

RTLDs provide a reliable, automated method for quantifying ICU registrar off-unit work patterns. This method may be adopted for quantifying other clinical staff work patterns in suitably equipped hospital environments.

References

1. National Outreach Forum: Operational Standards and Competencies for Critical Care Outreach Services. 2012; 20

2. Faculty of Intensive Care Medicine, Intensive Care Society: Guidelines for the Provision of Intensive Care Services 2016 Edition 1.1. 2016; 50–65

3. Nates JL, Nunnally M, Kleinpell R, et al.: ICU admission, discharge, and triage guidelines: A framework to enhance clinical operations, development of institutional policies, and further research. 2016.

4. Davies L, Davidoff F, Goodman D, et al.: SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence) . Am J Med Qual 2015; 30:543–549

Figures

Figure 1.

On-unit/off-unit analysis of ICU doctor location

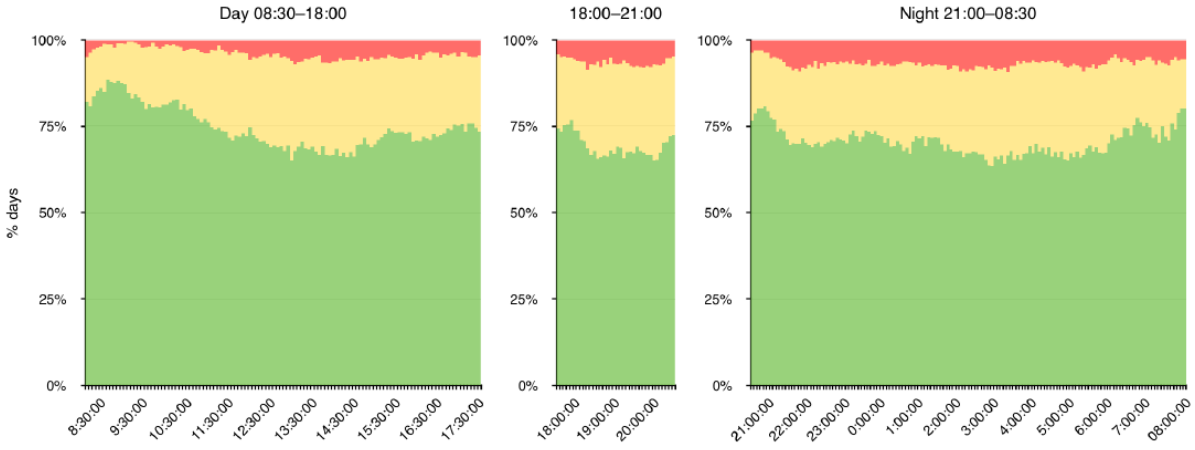
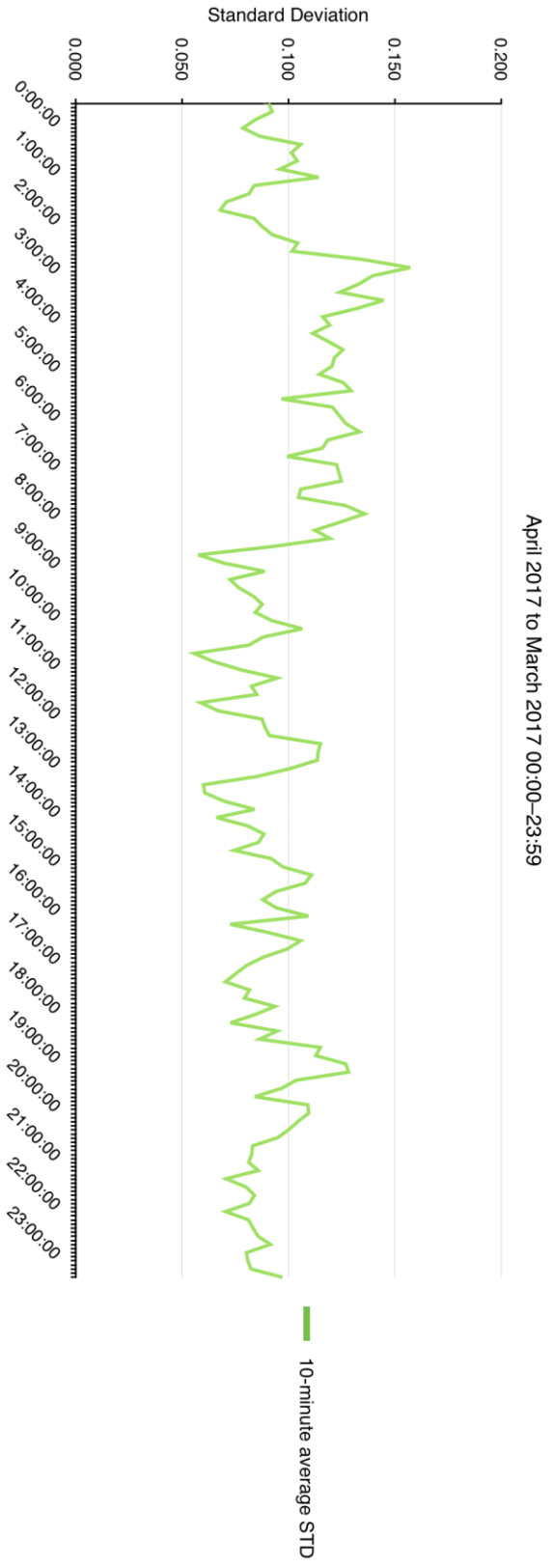


Figure 1: x-axis: 24h divided into five minutes increments y-axis: % of days (mean) that the five minute period had either none (red), one (amber) or both (green) doctors in the ICU. (Mean standard deviation is shown in the Appendix (Figure 1)).

Supplementary Material

Figure 1.

Variation (standard deviation) of the mean percentages of time spent on or off-unit.



The least variability is seen around handover times (0830)

Figure 2.

Tracker location variation between days of the week.

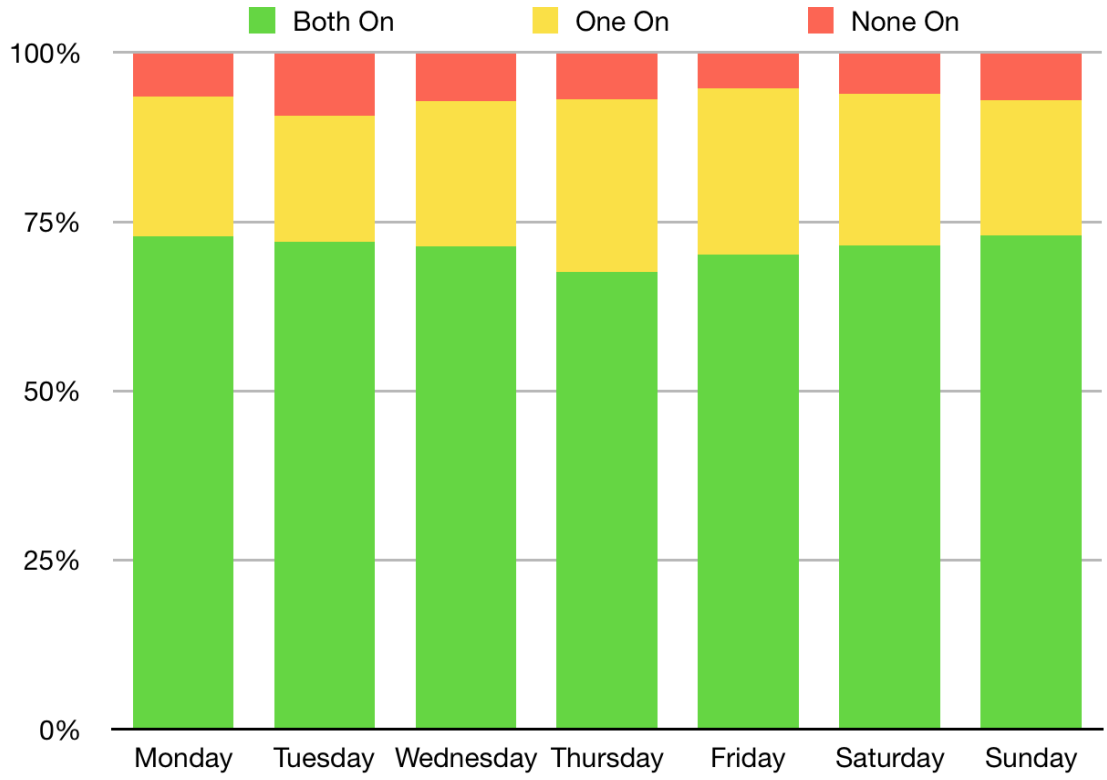


Figure 3a.

Tracker locations (total)

Tracker Locations April 2017–March 2018

	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
AICU	82%	82%	79%	79%	81%	80%	83%	84%	88%	92%	89%	86%	84%	82%	80%	83%	83%	85%	83%	81%	81%	86%	81%	81%
ED/EAU	9%	9%	9%	9%	8%	8%	6%	6%	5%	3%	5%	6%	7%	8%	9%	9%	9%	8%	7%	9%	11%	8%	11%	11%
Theatres	1%	1%	1%	1%	1%	1%	1%	1%	1%	0%	0%	0%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Radiology	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Trauma	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Changing Rooms	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Maternity	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Level 2	1%	0%	1%	1%	1%	1%	2%	1%	1%	1%	0%	1%	1%	2%	1%	1%	1%	1%	2%	1%	1%	1%	1%	1%
Level 4	1%	1%	1%	0%	0%	1%	0%	0%	0%	0%	0%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	0%	0%	1%
Level 5	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	1%	1%
Level 6	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	2%	1%
Level 7	1%	1%	1%	1%	1%	0%	1%	1%	1%	0%	0%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	0%	1%	1%
Level 8	0%	1%	3%	5%	5%	5%	3%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
West Wing	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	1%	1%	1%	1%	0%	0%	1%	0%	0%	1%	0%	0%
Other	1%	1%	2%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	2%	2%	1%	1%	1%	2%	2%	1%	1%	1%	1%
Null	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%

Figure 3b. *Tracker locations (as a % of time off-unit)*

Tracker Off-Unit Locations April 2017–March 2018

	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
ED/EAU	52%	52%	50%	53%	46%	46%	36%	35%	29%	16%	28%	36%	41%	44%	51%	52%	53%	44%	40%	52%	65%	45%	42%	62%
Theatres	4%	6%	4%	4%	7%	4%	7%	5%	4%	3%	2%	2%	3%	4%	8%	7%	5%	4%	6%	5%	3%	3%	2%	5%
Radiology	0%	1%	1%	1%	1%	1%	0%	1%	0%	1%	2%	3%	2%	3%	3%	2%	2%	2%	2%	2%	2%	0%	0%	1%
Trauma	2%	0%	0%	0%	1%	1%	0%	0%	0%	1%	1%	2%	1%	1%	1%	0%	0%	0%	0%	1%	1%	0%	1%	0%
Changing Rooms	0%	0%	1%	2%	2%	2%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Maternity	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Level 2	5%	2%	4%	4%	3%	6%	7%	11%	8%	6%	3%	5%	8%	11%	8%	7%	6%	7%	9%	8%	6%	4%	3%	6%
Level 4	4%	3%	3%	2%	2%	3%	3%	3%	2%	2%	2%	3%	4%	8%	6%	5%	5%	4%	5%	4%	3%	2%	1%	3%
Level 5	2%	3%	3%	1%	0%	2%	1%	2%	1%	1%	3%	2%	1%	2%	3%	1%	1%	3%	2%	2%	2%	2%	3%	3%
Level 6	7%	8%	7%	8%	5%	7%	6%	4%	5%	4%	4%	7%	8%	6%	6%	4%	3%	5%	6%	8%	5%	4%	6%	7%
Level 7	8%	4%	7%	4%	3%	3%	5%	8%	4%	3%	2%	4%	8%	5%	5%	3%	6%	4%	7%	8%	5%	2%	2%	5%
Level 8	2%	8%	20%	26%	27%	26%	19%	4%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%
West Wing	3%	1%	2%	1%	1%	1%	1%	3%	1%	1%	2%	3%	3%	4%	3%	4%	3%	2%	4%	3%	2%	3%	2%	2%
Other	5%	7%	10%	7%	5%	6%	6%	7%	4%	4%	6%	8%	8%	10%	11%	7%	8%	6%	9%	12%	8%	6%	5%	6%
Data Error	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	3%	5%

Figure 3a and 3b. Level 2: A mix of corridors, shops and outpatient departments, Levels 4 & 5: Specialty wards, Level 6: Surgical wards, Level 7: Medical wards, Level 8: the on-call room for the Tier 2 registrars. West Wing: Neurosurgical, Plastics and Ear, Nose and Throat Surgical Wards, Null/Data Error: 'Floor Hopped' Data.

Future Directions

Time and Motion Analysis

Time and motion analysis show the movement and activities of hospital staff as they deliver care to patients. It is plausible that hospital staff alter their movements and activities in response to patient demand (i.e., increase physical proximity because of the need to manage deterioration). These data may hypothetically be used independently, or be incorporated with existing models, to aid in the prediction of patient deterioration. To date this form of analysis has not occurred but steps have been taken towards developing methods that could enable it. Traditionally, data is collected manually by external observers, but recently automated computerised systems have become available. (1–5) Within the ICU, analysis has shown significant differences between the pattern of work in an ICU as compared to other medical and surgical units. (6) Likewise the impact of strain on ICU workflow may influence patient care. (7) ICU registrar activity outside the unit has been automatically quantified using Real Time Location Devices and may represent activity associated with deteriorating patients (Chapter 6). (8) Time and motion analysis has also been applied to patient journeys through the hospital. (9, 10)

Natural Language Processing

The analysis of free text in the EPR relies on natural language processing (NLP). NLP, also a form of Artificial Intelligence (AI), can model a range of clinically relevant data (e.g., symptoms, examination findings, clinical diagnosis), which may in turn be incorporated into models predicting clinical deterioration. This process combines two forms of AI: NLP and Machine Learning (ML). Korach et al. showed that unsupervised machine learning methods can successfully identify meaningful content in free-text nursing notes. (11) They demonstrated models using demographics and vital signs better predicted mortality when augmented by

nursing note NLP. (12) In ICU patients, NLP also augments mortality prediction model performance and prolonged length of stay prediction performance. (13–16) However, despite these promising findings NLP is yet to be implemented in real-time clinical workflows specific to deteriorating patients.

Patient Subgroups

Sepsis is a large and important subgroup of the deteriorating patient cohort. Two recent reviews showed there are more than 130 sepsis algorithms currently being tested. (17, 18) Both found methodological inconsistency and population heterogeneity and there remains a large gap between the creation of algorithms and their implementation into clinical practice. Further, in 2021 Wong et al. externally validated the Epic Sepsis Model (ESM) which has been implemented in hundreds of hospitals in the United States. ESM was found to have low sensitivity compared to current clinical practice, with low discrimination and calibration when predicting the onset of sepsis. (19)

COVID-19 has presented institutions with unique challenges, particularly with respect to the prediction and management of demand for ICU resources. AI has been applied in multiple domains with respect to COVID-19 including institution-level predictions, diagnosis, and prognostication. Systematic reviews of machine learning strategies applied to COVID-19 have been conducted previously. (20)(21) Noteworthy applications include the use of autoregressive models and time-delay artificial neural networks, in the prediction of case numbers and ICU bed availability. (22)(23) With respect to individual-level prediction of deterioration, models have been developed aiming to predict transfer to ICU, (24) intubation (25) and mortality.(26) The potential application of such models to the large numbers of individuals affected by COVID-19 highlights the importance of taking a deliberate approach aiming to use these models in an ethical manner that minimises bias and maximises public good. (27)

Additional Relevant Research

The high stakes involved in the clinical decision making around deteriorating patients emphasise the importance of having an ethical framework for the development and implementation of AI models. The potential for unintended ethical consequences is a concern in multiple areas, including bias, confidentiality and financial incentives. (28) Potential sources of bias include: biased datasets, models being developed that reflect human biases and automation bias adversely affecting human decision making. (29) Multiple strategies may be employed to help minimise these issues, including improving the interpretability of AI models and post-authorisation monitoring. (30) To mitigate these risks, models should be researched and employed with transparency to maintain and support public trust in medical AI applications. Finally, AI applications to wearable, remote vital sign monitoring equipment are expected to play an important role in future health care systems but remain experimental. (31)

Real world implementation

AI has revolutionised operations in personal banking, investment, manufacturing and news media but has not made comparable progress in health. (32) Despite the large increase in AI related clinical research over the last decade, very few algorithms and/or AI clinical applications progress through to the level of real-time clinical implementation. (33) This is demonstrated in the medical literature but also in FDA applications and approvals. (34) As of 2020, only 29 medical AI/ML specific applications were made and of these, none related to clinical deterioration (nearly all related to radiology or cardiology) (35) demonstrating that AI is still some way from being ubiquitous in health.

Research guidelines

The Consolidated Standards of Reporting Trial - Artificial Intelligence (CONSORT-AI) extension was released in 2020. (36) It was developed in parallel with its companion statement

SPIRIT-AI (both developed via consensus). (37) CONSORT-AI contains 14 new items that are specific to reporting AI research. These items include how the AI was integrated into the trial setting, data input and validity methodology, how the algorithm contributed to decision making, accessibility of algorithm code and algorithm performance error. CONSORT-AI is an important development in this field because research methodologies to date have been of varying quality and heterogenous. (34)

Conclusion

Given high numbers of patients continue to clinically deteriorate in hospital wards each year, optimising the systems around their care remains important. Developing HAVEN has been an important step but implementation into clinical workflows must occur to realise patient and organisational benefit. This work is currently underway in South Australia and in the United Kingdom. For the South Australian arm, a study protocol for an international retrospective validation study has been completed and has ethics approval. A prospective feasibility study and implementation study are currently being designed with ethics and funding applications in preparation.

Finally, a limitation of the project in Oxford was the frailty of the technical architecture for the analytics system. (38) To overcome similar limitations in South Australia, significant work has been undertaken to develop a cloud based platform (called HeartAI) to host HAVEN in its current and future forms (https://www.heartai.net//projects/haven-sa/index.html). It is anticipated this system will have enough capacity to host subsequent generations of HAVEN at scale and provide the technical foundation to drive this research forward.

References

1. Han B, Li Q, Chen X, et al.: Workflow for Intensive Care Unit Nurses: A Time and Motion Study [Internet]. *SAGE Open* 2020; 10:215824402094743Available from: http://dx.doi.org/10.1177/2158244020947433

2. Lopetegui M, Yen P-Y, Lai A, et al.: Time motion studies in healthcare: what are we talking about? *J Biomed Inform* 2014; 49:292–299

3. Li L, Hains I, Hordern T, et al.: What do ICU doctors do? A multisite time and motion study of the clinical work patterns of registrars. *Crit Care Resusc* 2015; 17:159–166

4. Wang X, Blumenthal HJ, Hoffman D, et al.: Modeling patient-related workload in the emergency department using electronic health record data. *Int J Med Inform* 2021; 150:104451

5. Benda NC, Blumenthal HJ, Hettinger AZ, et al.: Human Factors Design in the Clinical Environment: Development and Assessment of an Interface for Visualizing Emergency Medicine Clinician Workload. *IISE Transactions on Occupational Ergonomics and Human Factors* 2018; 6:225–237

6. Schenk E, Schleyer R, Jones CR, et al.: Time motion analysis of nursing work in ICU, telemetry and medical-surgical units. *J Nurs Manag* 2017; 25:640–646

7. Hefter Y, Madahar P, Eisen LA, et al.: A Time-Motion Study of ICU Workflow and the Impact of Strain. *Crit Care Med* 2016; 44:1482–1489

8. Malycha J, Murphy D, Barker G, et al.: Using real-time location devices (RTLD) to quantify off-unit adult intensive care registrar workload: a 1-year tertiary NHS hospital prospective observational study. *J Clin Monit Comput* 2019; 3–7

9. van Oostveen CJ, Gouma DJ, Bakker PJ, et al.: Quantifying the demand for hospital care services: a time and motion study. *BMC Health Serv Res* 2015; 15:15

10. Azzopardi M, Cauchi M, Cutajar K, et al.: A time and motion study of patients presenting at the accident and emergency department at Mater Dei Hospital [Internet]. *BMC Research Notes* 2011; 4Available from: http://dx.doi.org/10.1186/1756-0500-4-421

11. Korach ZT, Yang J, Rossetti SC, et al.: Mining clinical phrases from nursing notes to discover risk factors of patient deterioration. *Int J Med Inform* 2020; 135:104053

12. Korach ZT, Cato KD, Collins SA, et al.: Unsupervised Machine Learning of Topics Documented by Nurses about Hospitalized Patients Prior to a Rapid-Response Event. *Appl Clin Inform* 2019; 10:952–963

13. Marafino BJ, Boscardin WJ, Dudley RA: Efficient and sparse feature selection for biomedical text classification via the elastic net: Application to ICU risk stratification from nursing notes. *J Biomed Inform* 2015; 54:114–120

14. Marafino BJ, Park M, Davies JM, et al.: Validation of Prediction Models for Critical Care Outcomes Using Natural Language Processing of Electronic Health Record Data. *JAMA Netw Open* 2018; 1:e185097

15. Lehman L-W, Saeed M, Long W, et al.: Risk stratification of ICU patients using topic models inferred from unstructured progress notes. *AMIA Annu Symp Proc* 2012; 2012:505–511

16. Weissman GE, Hubbard RA, Ungar LH, et al.: Inclusion of Unstructured Clinical Text Improves

Early Prediction of Death or Prolonged ICU Stay. *Crit Care Med* 2018; 46:1125–1132

17. Fleuren LM, Klausch TLT, Zwager CL, et al.: Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* 2020; 46:383–400

18. Schinkel M, Paranjape K, Nannan Panday RS, et al.: Clinical applications of artificial intelligence in sepsis: A narrative review. *Comput Biol Med* 2019; 115:103488

19. Wong A, Otles E, Donnelly JP, et al.: External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med* 2021; 181:1065–1070

20. Wang L, Zhang Y, Wang D, et al.: Artificial Intelligence for COVID-19: A Systematic Review. *Front Med* 2021; 8:704256

21. Shakeel SM, Kumar NS, Madalli PP, et al.: COVID-19 prediction models: a systematic literature review. *Osong Public Health Res Perspect* 2021; 12:215–229

22. Goic M, Bozanic-Leal MS, Badal M, et al.: COVID-19: Short-term forecast of ICU beds in times of crisis. *PLoS One* 2021; 16:e0245272

23. Syrowatka A, Kuznetsova M, Alsubai A, et al.: Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. *NPJ Digit Med* 2021; 4:96

24. Cheng F-Y, Joshi H, Tandon P, et al.: Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients [Internet]. *J Clin Med Res* 2020; 9Available from: http://dx.doi.org/10.3390/jcm9061668

25. Arvind V, Kim JS, Cho BH, et al.: Development of a machine learning algorithm to predict intubation among hospitalized patients with COVID-19. *J Crit Care* 2021; 62:25–30

26. Mahdavi M, Choubdar H, Zabeh E, et al.: A machine learning based exploration of COVID-19 mortality risk. *PLoS One* 2021; 16:e0252384

27. Cave S, Whittlestone J, Nyrup R, et al.: Using AI ethically to tackle covid-19. *BMJ* 2021; 372:n364

28. Char DS, Shah NH, Magnus D: Implementing Machine Learning in Health Care - Addressing Ethical Challenges. *N Engl J Med* 2018; 378:981–983

29. Challen R, Denny J, Pitt M, et al.: Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019; 28:231–237

30. Vokinger KN, Feuerriegel S, Kesselheim AS: Mitigating bias in machine learning for medicine. *Commun Med* 2021; 1:25

31. Nahavandi D, Alizadehsani R, Khosravi A, et al.: Application of artificial intelligence in wearable devices: Opportunities and challenges. *Comput Methods Programs Biomed* 2022; 213:106541

32. Martin G Seneviratne Nigam H Shah Larry Chu: Bridging the implementation gap of machine learning in healthcare [Internet]. *BMJ Innovation* 2020; Available from: http://dx.doi.org/bmjinnov-2019-000359

33. Meskó B, Görög M: A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med* 2020; 3:126

34. Mann KD, Good NM, Fatehi F, et al.: Predicting Patient Deterioration: A Review of Tools in the Digital Hospital Setting. *J Med Internet Res* 2021; 23:e28209

35. Benjamens S, Dhunnoo P, Meskó B: The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med* 2020; 3:118

36. Liu X, Cruz Rivera S, Moher D, et al.: Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020; 26:1364–1374

37. Cruz Rivera S, Liu X, Chan A-W, et al.: Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* 2020; 26:1351–1363

38. Dahella SS, Briggs JS, Coombes P, et al.: Implementing a system for the real-time risk assessment of patients considered for intensive care. 2020; 0:1–7

Appendix A

Presentations at national and international meetings

The student presented studies at conferences run by intensive care medicine or rapid response system societies completed during his doctoral programme as oral or poster presentations.

National Presentations

Australia and New Zealand Intensive Care Society, Clinical Trials Group. Meeting. 2021
Hospital Alerts Via Electronic Noticeboard (HAVEN) – Australian Update.
Noosa, Australia

Australia and New Zealand Intensive Care Society, Clinical Trials Group. Meeting. 2020
Data driven improvements in predicting clinical deterioration.
Noosa, Australia

International Presentations

Society of Rapid Response Systems Annual Conference 2018
The effect of Fractional Inspired Oxygen Concentration on EWS Performance
Manchester, UK

State of the Art Meeting 2017-
Real time location tracking of adult intensive care unit (ICU) registrars in a NHS hospital.
Liverpool, UK

Appendix B

Prizes awarded during candidature

2019

University of Oxford, Nuffield Department of Clinical Neurosciences

Runner-Up for the Thomas Willis Early Career Research Prize

2018

The International Society of Rapid Response Systems and The National Outreach Forum, Winner for the best abstract

2018

Oxford Regional Intensive Care Society Spring Meeting

Winner for the best poster for foundation doctor (supervisor)

Appendix C

Grants awarded during candidature

Ludbrook G, Grocott M, Bogdon S, Malycha J. **Advanced Recovery Room Care – an iterative model to improve outcomes and reduce cost in perioperative care.** 2020 Central Adelaide Health Local Health Network Clinical Rapid Implementation Project Scheme (CRIPS)
$200,000

Appendix D

Other publications during candidature

2021

Wronikowska M, Malycha J, Morgan L, Westgate V, Petrinic T, Young JD, Watkinson PJ. **Systemic review of applied usability metrics within usability evaluation methods for hospital electronic healthcare record systems.** *Journal of Evaluation in Clinical Practice*. Accepted for publication May 2021 (Manuscript ID JECP-2020-0823).

2019

C Subbe, J Bannard-Smith, J Bunch, R Champunot, MA DeVita, L Durham, DP Edelson, I Gonzalez, C Hancock, R Haniffa, J Hartin, H Haskell, H Hogan, DA Jones, CJ Kalkman, GK Lighthall, J Malycha, AV Phillips, F Rubulotta, RK So, J Welch. **Quality Metrics for the Evaluation of Rapid Response Systems: Proceedings from the third international consensus conference on Rapid Response Systems**. *Resuscitation* 2019; 141: 1-12.

Smith GB, Redfern OC, Pimentel MA, Gerry S, Collins GS, Malycha J, Prytherch D, Schmidt PE, Watkinson, PJ. **The National Early Warning Score 2 (NEWS2).** *Clinical Medicine* 2019; No 3: 260-3.

Pimentel MA, Smith GB, Redfern OC, Gerry S, Collins GS, Malycha J, Prytherch D, Schmidt PE, Watkinson PJ. **Reply to: NEWS2 needs to be tested in prospective trials involving patients with confirmed hypercapnia**. *Resuscitation* 2019; 139: 371-372.

Pimentel M, Redfern O, Malycha J, Collins G, Gerry S, Prytherch D, Schmidt P, Smith G, Watkinson P. **An evaluation of the ability of the National Early Warning Score 2 and Chronic Respiratory Early Warning Score (CREWS) modifications to the National Early Warning Score to identify patients at risk of in-hospital mortality: a multicentre database study**. *Resuscitation* 2019; 134: 147-156.