



University of Southern Denmark

A tale of two continents

The performance of six early warning scores in two emergency departments

Lo, Ronson Sze Long; Lyngholm, Le Elias; Brabrand, Mikkel; Leung, Ling Yan; Hung, Kevin Kei Ching; Laugesen, Sune; Posth, Stefan; Cooksley, Tim; Graham, Colin A.; Nickel, Christian H.

Published in:

Hong Kong Journal of Emergency Medicine

DOI:

10.1177/10249079231153719

Publication date:

2023

Document version:

Final published version

Document license:

CC BY-NC

Citation for pulished version (APA):

Lo, R. S. L., Lyngholm, L. E., Brabrand, M., Leung, L. Y., Hung, K. K. C., Laugesen, S., Posth, S., Cooksley, T., Graham, C. A., & Nickel, C. H. (2023). A tale of two continents: The performance of six early warning scores in two emergency departments. *Hong Kong Journal of Emergency Medicine*, 30(6), 374-380. <https://doi.org/10.1177/10249079231153719>

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use

This work is brought to you by the University of Southern Denmark.

Unless otherwise specified it has been shared according to the terms for self-archiving.

If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk



A tale of two continents: The performance of six early warning scores in two emergency departments

Hong Kong Journal of Emergency Medicine
1–7

© The Author(s) 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/10249079231153719

journals.sagepub.com/home/hkj



Ronson Sze Long Lo¹ , Le Elias Lyngholm²,
Mikkel Brabrand^{2,3,4} , Ling Yan Leung¹,
Kevin Kei Ching Hung¹ , Sune Laugesen³, Stefan Posth³,
Tim Cooksley⁵, Colin A Graham¹ and Christian H Nickel⁶

Abstract

Background: Several early warning scores have been designed to optimize acute care by identifying patients at risk of deterioration.

Methods: In this post hoc dual center study, we analyzed the performance of six clinical scores (the Goodacre score, Groarke, Worthing Physiological Score, Rapid Acute Physiology Score, Rapid Emergency Medicine Score, United Kingdom National Early Warning Score). The primary outcome is 30-day all-cause mortality after inclusion and data were obtained from previous studies performed at two different emergency departments on two continents (Denmark, Europe, and Hong Kong, Asia).

Results: We included 2952 people; 1482 (50.2%) were male, mean age (standard deviation) was 65.7 (18.3) years, and 109 (3.7%) died within 30 days. Mortality rate increased steadily with increasing scores for all six scoring systems in Hong Kong while this was less obvious in Denmark. In all patients, Rapid Acute Physiology Score had the lowest discriminatory power while National Early Warning Score had the highest. National Early Warning Score performed best in Hong Kong while Worthing performed marginally better in Denmark.

Discussion: Surprisingly, the performance of the scoring systems varied considerably, but were largely unaffected by location, and none of them performed close to what clinicians would normally require for predicting 30-day all-cause mortality

Conclusion: All scores performed similarly across both centers, with poor prediction of 30-day all-cause mortality. Based on these findings, we believe that clinical scores must be supplemented by either biochemical values or global markers of physiological reserve to reflect reality and to be of true value.

Keywords

Goodacre score, Groarke, Rapid Acute Physiology Score, risk prediction, United Kingdom National Early Warning Score

¹Accident and Emergency Medicine Academic Unit, Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China

²Department of Emergency Medicine, Hospital of South West Jutland, Esbjerg, Denmark

³Department of Emergency Medicine, Odense University Hospital, Odense, Denmark

⁴Institute of Regional Health Research, University of Southern Denmark, Esbjerg, Denmark

⁵Manchester University NHS Foundation Trust, Manchester, UK

⁶Emergency Department, University Hospital Basel, Basel, Switzerland

Corresponding author:

Kevin Kei Ching Hung, Accident and Emergency Medicine Academic Unit, Chinese University of Hong Kong, Prince of Wales Hospital, 2/F, Main Clinical Block and Trauma Centre, Shatin, New Territories, Hong Kong, China.

Email: kevin.hung@cuhk.edu.hk



Introduction

There is an abundance of early warning scores^{1,2} designed to identify patients at risk of deterioration so that acute care can be optimized.¹ A substantial number of scores have undergone local validation and many international scrutiny.² Some scores have demonstrated utility while others are less than optimal.²

While most scores contain identical physiological variables, the assignments of their weightings vary. While one early warning score assigns a score of 1 for a systolic blood pressure of 70 mmHg, another will opt for a score of 3, affecting the performance of the score.³

Our group has recently published a study from Denmark validating six clinical scores.⁴ Apart from being single centered, this data set was also affected by missing data. To rectify these shortcomings, we have designed this post hoc study to analyze already collected data from two very different emergency departments (EDs) placed on two different continents to test the performance of the six clinical scores (the Goodacre score,⁵ Groarke score,⁶ Worthing Physiological score,⁷ Rapid Acute Physiology Score (RAPS),⁸ Rapid Emergency Medicine Score (REMS),⁹ and the United Kingdom National Early Warning Score (NEWS)¹) anew.

Methods

We performed a post hoc dual center analysis of already collected data from two EDs, one in Denmark and one in Hong Kong, to examine the performance of six early warning scores in common clinical use. Both cohorts were collected with another aim in mind (ClinicalTrials.gov Identifier: NCT03108807 and NCT02817581).

Settings

The Danish cohort was collected from the xxx, a 400-bed regional teaching hospital where all patients except women in labor and children are admitted through the ED. The cohort from Hong Kong was collected at the ED of the xxx, a 1500-bed tertiary university hospital with an open access ED. The Danish department sees approximately 40 patients per day, while the Hong Kong ED deals with approximately 400 patients per day.

At both centers, all adult (age ≥ 18 years) patients, who were not in the highest triage level or who did not have obviously minor problems (i.e. the lowest triage level) at their first visit in the study period, were included in this analysis. Patients were not necessarily hospitalized after inclusion and some were discharged directly from the ED.

Scores

We included six scores^{1,5-9} that mostly were similar, but had subtle differences (Table 1). All scores were based on physiological parameters, but the weights varied.

Outcome

Our primary outcome is 30-day all-cause mortality after inclusion. In Denmark, follow-up was conducted through the Danish Civil Personal Register,¹⁰ while follow-up in Hong Kong was via telephone follow-up calls and using the local electronic healthcare system (Clinical Management System). Blinding was not performed.

Sample size

As this is a post hoc analysis, a sample size calculation has not been performed.

Statistical analysis

Data are presented descriptively as mean (standard deviation (SD)) or number (proportion) as appropriate. The discriminatory power, that is, the ability to discriminate between patients that meet the endpoint, is presented as area under the receiver operating characteristics curve (AUROC). Calibration, that is, the precision, was calculated according to Seymour et al.¹¹ Analyses were performed using Stata 15 (Stata Corp, College Station, TX USA).

Results

We included 2952 people; 1482 (50.2%) were male and the mean age (SD) was 65.7 (18.3) years (Table 2). Within 30 days, 109 (3.7%) died, the majority in Hong Kong (Table 2).

The mortality rate increased steadily with increases in scores for all six scores in Hong Kong (Figure 1(a)), while this was less obvious in Denmark (Figure 1(b)).

The discriminatory power varied between the six scores. In all patients, RAPS had the lowest AUROC of 0.547, while NEWS had the highest at 0.701 (Table 3 and Figure 2). NEWS also had the best discriminatory power in Hong Kong, while Worthing performed marginally better in Denmark. All scores had acceptable calibration in both settings (Table 3).

Discussion

In this post hoc study from two EDs on two continents, we found that the performance of six early warning scores in common clinical use varied considerably, but not due to the location. While there were minor differences between the sites, all scores had very similar performances across both sites.

While all scores had a similar predictive power in an undifferentiated ED population, none of them had a performance close to what clinicians would normally require for predicting 30-day all-cause mortality.² Obviously, some of this is explained by the fact that the scores were developed for predicting short-term mortality, but we believe, as we have previously argued,¹² that scores based on purely vital

Table 1. Physiological parameters included in each of the scores.

	Respiratory rate	Oxygen saturation	Temperature	Blood pressure	Heart rate	Consciousness	Age
NEWS	✓	✓	✓	✓	✓	✓	
RAPS	✓			✓	✓	✓	
REMS	✓	✓	✓	✓	✓	✓	✓
Worthing	✓	✓	✓	✓	✓	✓	
Groarke	✓		✓	✓	✓	✓	
Goodacre		✓				✓	✓

NEWS: National Early Warning Score; RAPS: Rapid Acute Physiology Score; REMS: Rapid Emergency Medicine Score.

Table 2. Baseline characteristics of the participants.

	Hong Kong	Denmark	All
Number (%)	1253 (42.5)	1697 (57.5)	2952 (100.0)
Male, number (%)	638 (50.8)	844 (49.7)	1,482 (50.2)
Age (mean), years (SD)	69.5 (17.8)	62.9 (18.2)	65.7 (18.3)
Length of stay, days (SD)	6.8 (6.5)	2.9 (7.1)	—
30-day mortality, number (%)	71 (5.7)	38 (2.2)	109 (3.7)
Missing NEWS, number (%)	46 (3.7)	101 (6.0)	147 (5.0)
Missing RAPS, number (%)	44 (3.5)	81 (4.8)	125 (4.2)
Missing REMS, number (%)	46 (3.7)	102 (6.0)	148 (5.0)
Missing Worthing, number (%)	46 (3.7)	101 (6.0)	147 (5.0)
Missing Groarke, number (%)	45 (3.6)	98 (5.8)	143 (4.8)
Missing Goodacre, number (%)	4 (0.3)	51 (3.0)	55 (1.9)

SD: standard deviation; NEWS: National Early Warning Score; RAPS: Rapid Acute Physiology Score; REMS: Rapid Emergency Medicine Score.

signs are likely unable to stand on their own. They need to be supplemented by either biochemical values, such as D-dimer,² or global markers of physiological reserve, such as mobility^{13,14} to improve their performance and support clinical decision-making in an emergency setting.

The components of the six scores varied but the main difference was in the weights assigned to the abnormal values. REMS⁹ has the most components (seven), whereas the Goodacre score⁵ has three and RAPS⁸ has four. The scores with the most components generally performed better. NEWS¹ and Worthing⁷ performed well at both sites and we believe the difference in their predictive power to be due to the weights. Interestingly though, the Goodacre score,⁵ with its only three components, also performed well. This could indicate that identifying patients at risk perhaps is not only based on strict measurement of all vital signs but discerning the key predictive variables.

The Goodacre score⁵ contains few traditional vital signs and is similar to components of the Emergency Severity Index (ESI) triage system.¹⁵ It only comprises age, consciousness, and oxygen saturation. Chronological age does give an indication of chronic health status and the potential co-morbidity burden while decreased consciousness can be considered a universal marker for a severe deterioration in health.¹⁶

It is somewhat surprising that all six scores performed equally in both settings. The healthcare systems in Denmark and Hong Kong are different and the populations attending the ED are difficult to compare. While Denmark has a very strong primary care system, Hong Kong has a limited primary care system. In Denmark, patients are either seen by a general practitioner before coming to the ED, or they attend after having contacted the EMS because of a suspected acute, life-threatening situation. In Hong Kong, anyone can visit the ED and register to be seen by an ED doctor. The reason for this surprising result is probably a difference in populations included in the cohorts. In Denmark, we actively sought inclusion of all patients arriving to the ED, while in Hong Kong, we actively sought inclusion of intermediate risk patients and thus excluded a lower risk population. But indeed, our group has previously shown that patients in Denmark and Uganda with near normal vital signs have comparable outcomes.¹⁷

Our study has some limitations. While we tried to keep rates of missing data as low as possible, we still had up to 6% missing values. In addition, none of the scores were designed to predict clinical outcomes 30 days into the future. Any prediction system will become increasingly inaccurate as it tries to reach further into the future. As the vital signs were mostly collected by clinical staff and then

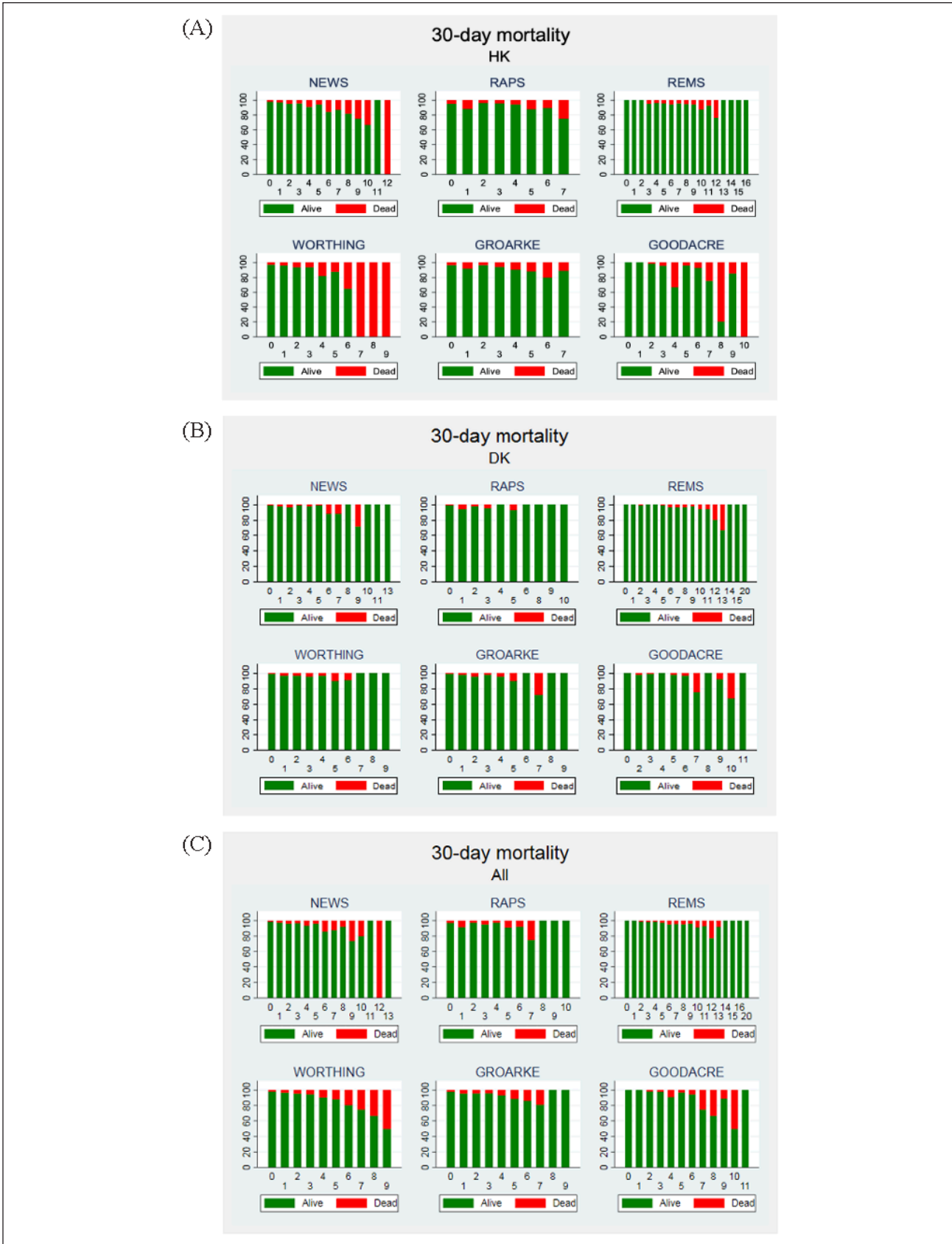


Figure 1. Bar graphs of 30-day mortality rate with increasing scores for patients from (a) Hong Kong, (b) Denmark, and (c) all. The green color designates survivors while the red color shows fatalities.

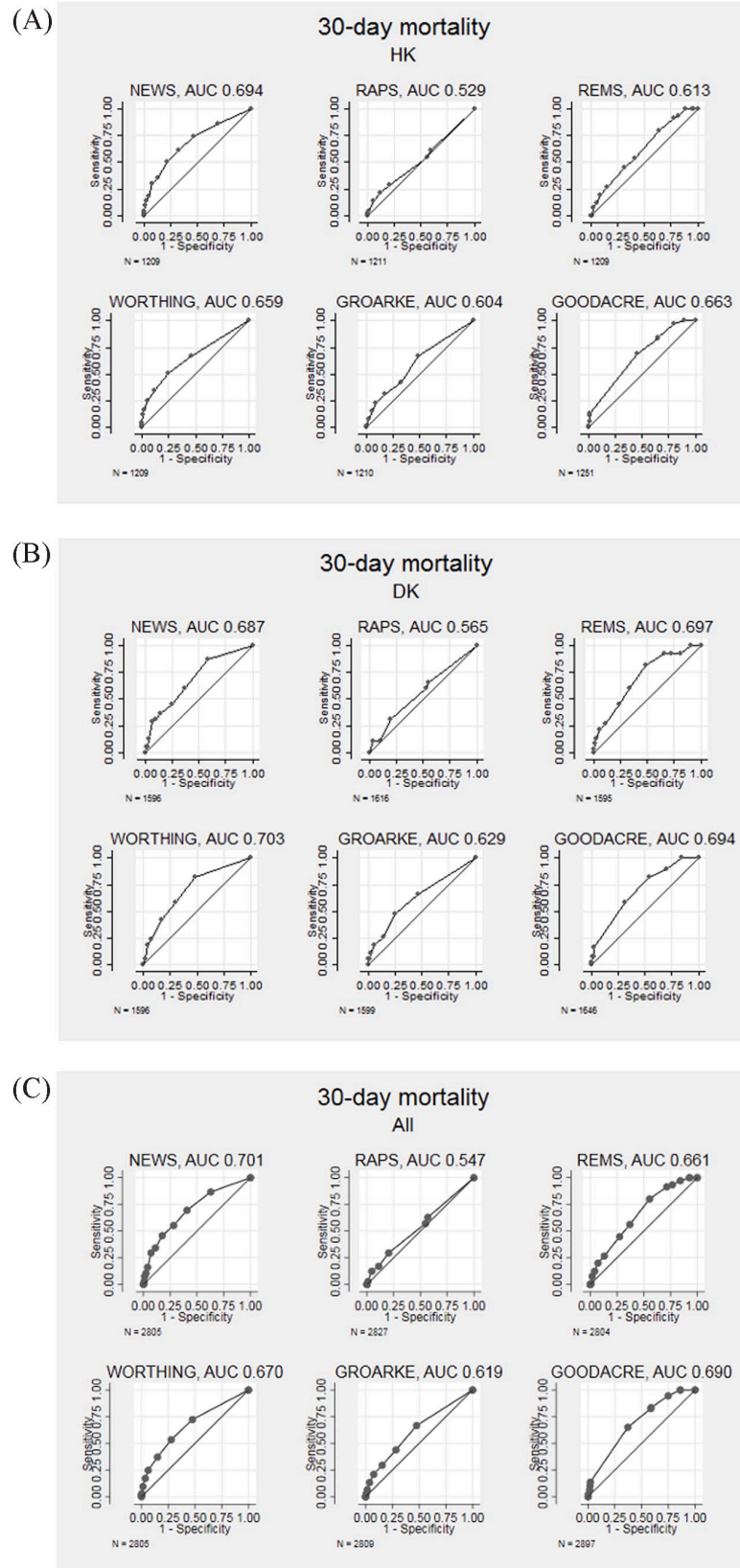


Figure 2. Discrimination plots for the six clinical scores, used in (a) Hong Kong, (b) Denmark, and (c) all.

Table 3. Discriminatory power and calibration of the six scores.

	Hong Kong		Denmark		All	
	Discrimination	Calibration (<i>p</i>)	Discrimination	Calibration (<i>p</i>)	Discrimination	Calibration (<i>p</i>)
NEWS	0.695 (0.627–0.763)	0.81	0.687 (0.604–0.769)	0.14	0.701 (0.649–0.752)	0.41
RAPS	0.529 (0.456–0.602)	0.30	0.565 (0.475–0.656)	0.08	0.547 (0.490–0.603)	0.12
REMS	0.613 (0.548–0.678)	0.48	0.697 (0.619–0.775)	0.53	0.661 (0.613–0.709)	0.45
Worthing	0.659 (0.587–0.731)	0.61	0.703 (0.622–0.785)	0.36	0.671 (0.616–0.725)	0.76
Groarke	0.604 (0.535–0.672)	0.10	0.629 (0.538–0.721)	0.54	0.619 (0.565–0.674)	0.38
Goodacre	0.663 (0.606–0.719)	0.08	0.694 (0.617–0.771)	0.35	0.690 (0.646–0.734)	0.13

NEWS: National Early Warning Score; RAPS: Rapid Acute Physiology Score; REMS: Rapid Emergency Medicine Score.

copied by the research staff, we cannot rule out inaccurate measurements or errors in data entry.

Conclusion

In this post hoc study from two EDs in Denmark and Hong Kong, we found that six early warning scores in common clinical use varied in performance as measured using discriminatory power and calibration. All scores performed equally across both sites. The best scores were NEWS and Worthing Score, while RAPS had the poorest performance. Surprisingly, the Goodacre score with only three ingredients (age, consciousness, and oxygen saturation) performed almost as well as the best score.

Acknowledgements

The authors wish to thank Dr John Kellett for his invaluable help with writing the manuscript.

Author contributions

M.B., L.Y.L., K.K.C.H., C.A.G., and C.H.N. conceived the study. M.B., C.A.G., T.C., and C.H.N. were involved in protocol development. R.S.L.L. wrote the initial article. L.E.L., R.S.L.L., L.Y.L., K.K.C.H., S.L., S.P., and T.C. collected and analyzed the data. All authors approved the final version of the article and edited and reviewed the article.

Availability of data and materials

The Hong Kong data set is available from the corresponding author. The Danish data set cannot be shared due to Danish law.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Informed consent

Written consent was obtained either from the patient or from a relative in all cases.

Ethical approval

The Danish part of the study was approved by Danish Regional Committee of Health Research Ethics (Identifier: S-20170005) and the Danish Data Protection Agency (Identifier: Region Syddanmark 2452). The data from the Hong Kong cohort were obtained from a prospective study that was approved by the Institutional Review Board of The Joint Chinese University of Hong Kong—New Territories East Cluster Clinical Research Ethics Committee (CRE-2016.236).

ORCID iDs

Ronson Sze Long Lo  <https://orcid.org/0000-0003-3822-4918>
 Mikkel Brabrand  <https://orcid.org/0000-0002-3340-8251>
 Kevin Kei Ching Hung  <https://orcid.org/0000-0001-8706-7758>
 Colin A Graham  <https://orcid.org/0000-0002-4381-7470>

References

- 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(4): 465–470.
- Brabrand M, Folkestad L, Clausen NG, et al. Risk scoring systems for adults admitted to the emergency department: a systematic review. *Scand J Trauma Resusc Emerg Med* 2010; 18(1): 8.
- Jarvis S, Kovacs C, Briggs J, et al. Can binary early warning scores perform as well as standard early warning scores for discriminating a patient's risk of cardiac arrest, death or unanticipated intensive care unit admission? *Resuscitation* 2015; 93: 46–52.
- Brabrand M, Hallas P, Hansen SN, et al. Using scores to identify patients at risk of short term mortality at arrival to the acute medical unit: a validation study of six existing scores. *Eur J Intern Med* 2017; 45: 32–36.
- Goodacre S, Turner J and Nicholl J. Prediction of mortality among emergency medical admissions. *Emerg Med J* 2006; 23(5): 372–375.
- Groarke J, Gallagher J, Stack J, et al. Use of an admission early warning score to predict patient morbidity and mortality and treatment success. *Emerg Med J* 2008; 25(12): 803–806.
- Duckitt R, Buxton-Thomas R, Walker J, et al. Worthing physiological scoring system: derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study. *Br J Anaesth* 2007; 98(6): 769–774.

8. Rhee KJ, Fisher CJ Jr and Willitis NH. The rapid acute physiology score. *Am J Emerg Med* 1987; 5(4): 278–282.
9. Olsson T, Terent A and Lind L. Rapid Emergency Medicine Score can predict long-term mortality in nonsurgical emergency department patients. *Acad Emerg Med* 2004; 11(10): 1008–1013.
10. Schmidt M, Pedersen L and Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol* 2014; 29(8): 541–549.
11. Seymour CW, Kahn JM, Cooke CR, et al. Prediction of critical illness during out-of-hospital emergency care. *JAMA* 2010; 304(7): 747–754.
12. Brabrand M and Kellett J. Mobility measures should be added to the National Early Warning Score (NEWS). *Resuscitation* 2014; 85(9): e151.
13. Brabrand M, Kellett J, Opio M, et al. Should impaired mobility on presentation be a vital sign? *Acta Anaesthesiol Scand* 2018; 62(7): 945–952.
14. Nickel CH, Kellett J, Ortega RN, et al. Mobility identifies acutely ill patients at low risk of in-hospital mortality: a prospective multicenter study. *Chest* 2019; 156(2): 316–322.
15. Mistry B, De Ramirez SS, Kelen G, et al. Accuracy and reliability of emergency department triage using the emergency severity index: an international multicenter assessment. *Ann Emerg Med* 2018; 71(5): 581.e3–587.e3.
16. Bech CN, Brabrand M, Mikkelsen S, et al. Risk factors associated with short term mortality changes over time, after arrival to the emergency department. *Scand J Trauma Resusc Emerg Med* 2018; 26(1): 29.
17. Nabayigga B, Kellett J, Brabrand M, et al. The mortality of acutely ill medical patients for up to 60 days after admission to a resource poor hospital in sub-Saharan Africa compared with patients of similar illness severity admitted to a Danish Regional Teaching Hospital—an exploratory observational study. *Eur J Intern Med* 2016; 27: 24–30.