BMJ Open

Performance of digital early warning score (NEWS2) in a cardiac specialist setting: retrospective cohort study

Baneen Alhmoud ,[1] Tim Bonnici,[1,2] Daniel Melley,[3] Riyaz Patel ,[1,2] Amitava Banerjee [1,2]

To cite: Alhmoud B, Bonnici T, Melley D, *et al*. Performance of digital early warning score (NEWS2) in a cardiac specialist setting: retrospective cohort study. *BMJ Open* 2023;**13**:e066131. doi:10.1136/bmjopen-2022-066131

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-066131).

Received 29 June 2022
Accepted 23 February 2023

Check for updates

[1]University College London, London, UK
[2]University College London Hospitals NHS Foundation Trust, London, UK
[3]Barts Health NHS Trust, London, UK

Correspondence to
Professor Amitava Banerjee; ami.banerjee@ucl.ac.uk

ABSTRACT

Introduction Patients with cardiovascular diseases (CVD) are at significant risk of developing critical events. Early warning scores (EWS) are recommended for early recognition of deteriorating patients, yet their performance has been poorly studied in cardiac care settings. Standardisation and integrated National Early Warning Score 2 (NEWS2) in electronic health records (EHRs) are recommended yet have not been evaluated in specialist settings.

Objective To investigate the performance of digital NEWS2 in predicting critical events: death, intensive care unit (ICU) admission, cardiac arrest and medical emergencies.

Methods Retrospective cohort analysis.

Study cohort Individuals admitted with CVD diagnoses in 2020; including patients with COVID-19 due to conducting the study during the COVID-19 pandemic.

Measures We tested the ability of NEWS2 in predicting the three critical outcomes from admission and within 24 hours before the event. NEWS2 was supplemented with age and cardiac rhythm and investigated. We used logistic regression analysis with the area under the receiver operating characteristic curve (AUC) to measure discrimination.

Results In 6143 patients admitted under cardiac specialties, NEWS2 showed moderate to low predictive accuracy of traditionally examined outcomes: death, ICU admission, cardiac arrest and medical emergency (AUC: 0.63, 0.56, 0.70 and 0.63, respectively). Supplemented NEWS2 with age showed no improvement while age and cardiac rhythm improved discrimination (AUC: 0.75, 0.84, 0.95 and 0.94, respectively). Improved performance was found of NEWS2 with age for COVID-19 cases (AUC: 0.96, 0.70, 0.87 and 0.88, respectively).

Conclusion The performance of NEWS2 in patients with CVD is suboptimal, and fair for patients with CVD with COVID-19 to predict deterioration. Adjustment with variables that strongly correlate with critical cardiovascular outcomes, that is, cardiac rhythm, can improve the model. There is a need to define critical endpoints, engagement with clinical experts in development and further validation and implementation studies of EHR-integrated EWS in cardiac specialist settings.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Our study is the first to examine the performance of universal early warning score (EWS) (National Early Warning Score 2, NEWS2) in patients with cardiovascular diseases in a cardiac specialist hospital.
⇒ We were able to extract data from electronic health record systems where NEWS2 is integrated and automated, reflecting the accuracy of captured parameters and enabled us to integrate other data sources for critical outcomes and COVID-19 cases.
⇒ The study followed a retrospective data collection from three data sources where there was less control of missingness of NEWS2 recordings, heart rhythm at several points in time and other parameters that could be examined like FiO_2 level.
⇒ We conducted external validation of NEWS2; with our hospital data and internally; with the same data set used for supplementing the model with other parameters—external validation studies are needed for generalisability.
⇒ The endpoints examined were favoured by researchers in validation studies but may not be the ideal or precise points to measure triggered EWS against.

INTRODUCTION

Disease severity classification of patients with cardiovascular diseases (CVD) is challenging for nurses and physicians. Individuals with CVD can present with various sorts of critical events due to the disease's pathophysiology or the comorbidities associated.[1] The aetiology of the disease and the specialised care provided may impose the standardised use of deterioration risk scores, such as the widely adopted National Early Warning Score 2 (NEWS2).[2] Heterogeneity in the process of patient deterioration complicates detection and escalation.

CVD is the leading cause of death in the UK and worldwide, with an estimated healthcare cost of £9 billion annually.[1 3] Short-term critical events, such as cardiac arrest and transfer to the intensive care unit (ICU), are common in patients with CVD.[4 5] In addition, mortality and morbidity are major concerns in patients with CVD globally.[1] Risk stratification tools for long-term outcomes have long been favoured in this disease subgroup. Models like the Global Registry of Acute Coronary Events (GRACE) and the CHADS2-VASc are validated long-term risk scores in CVDs.[6 7]

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

Figure 1 The National Early Warning Score 2 (NEWS2).
CVPU, new Confusion, Voice, Pain, Unresponsive; SpO2, peripheral capillary oxygen saturation.

For early-stage risk prediction, risk stratification of critical deterioration was unified for all disease groups and settings using early warning scoring (EWS) systems.[8,9]

The recently implemented and developed NEWS2 has been recommended by the consensus of clinical experts. A minimum set of physiological parameters and standardised application across the National Health Service (NHS) to promote patient safety and unify clinicians' practice were the aims of NEWS2.[9–11] The predictability of NEWS and NEWS2 of critical events was found fair to be acceptable to the emergency department (ED)[12,13] and medical and surgical settings. On the other hand, issues reported in specialised settings, like poor predictive performance in haematology settings[13] and the need to supplement NEWS in emergency settings[12,14] and for patients with COVID-19,[9] indicate its inefficiency in particular contexts. In cardiac settings, studies on EWS performance in critical events are poor and insufficient to defend the application for escalating the acutely ill. A study in 2012 showed high predictive ability in cardiac patients using ViEWS (VitalpacTM-derived early warning score), a model with different parameters than NEWS.[15] Another examined RACE (RApid Clinical Evaluation), a postoperative cardiac EWS tool for cardiac surgical ICU; a narrowly included subset of cardiac patients.[16] Little is known about the predictive value of some NEWS2 components to be deemed reliable in this subgroup, that is, inclusion of temperature or missing heart rhythm (figure 1). Despite the need for reliable early deterioration detection, it is reported that specialised cardiac centres may overlook the value of developing and validating EWS. The complexity of models and difficulties in analysing electronic health record (EHR) data formed barriers to the validation of EWS.[16] However, in the era of predictive modelling generated from EHRs, and EWS embedded in EHR, validating EWS in patient population with a high rate of critical events is necessary. In the cardiac setting examined, NEWS2 was integrated into patients' EHRs, providing a quick and reliable method of calculation and scoring. In addition, automated monitoring was implemented to facilitate the routine monitoring recording into a ready-to-calculate and evaluate NEWS2. This digitalisation process has given a unique opportunity for

Table 1 Terms and definitions

Term	Definition
NEWS2	The National Early Warning Score 2: a model designed to perform as a tool for standardised assessment of acute illness severity by assessing physiological parameters and producing a score that helps in stratifying the risk of developing a worsening event.[29]
Supplementing a model	The process of combining variables with the original model, that were not part of it, to improve its discriminatory ability.[18,30]
Discrimination:	The ability of a model to distinguish between the patients who will develop an outcome of interest and the ones who will not.[31]
Automated monitoring	Integration of patients' routine monitoring and EHRs by transmitting measurements directly from monitoring machine to patients' charts and continuously calculating and updating NEWS2.[32,33]

EHR, electronic health record.

digital NEWS2 validation in a poorly studied subgroup with high incidents of critical events (table 1).

Objective

To investigate the performance of digital NEWS2 in predicting critical events, at admission and prior to deterioration, for cardiac patients in the COVID-19 context in a cardiac specialist setting.

Our specific aims are:

- To explore the independent association of physiological parameters and NEWS2 at hospital admission and 24 hours prior critical events; with disease severity (ICU admission, cardiac arrest, medical emergency and death).
- To examine the predictive ability of NEWS2 and the supplemented NEWS2 with potential determinants of disease severity on admission and 24 hours prior to the condition worsening.
- To compare the predictive value of NEWS2 with supplemented NEWS2 models (table 1).

METHODS

Study cohort

The study population was defined as adult patients admitted to St Bartholomew's Hospital, a cardiac specialist and teaching hospital in London with heart and cancer centres, from January to December 2020, under cardiac specialty care; for more than 24 hours. Due to conducting the study during the time nature of the pandemic, we have identified patients with CVD with COVID-19 based on positive PCR test results on or during admission.

Measures

NEWS2 and physiological parameters

We included physiological parameters and NEWS2 scores routinely obtained at hospital admission and 24 hours prior to deterioration. Included parameters that form NEWS2 score (figure 1) were respiratory rate (breaths per minute), oxygen saturation (%), systolic blood pressure (SBP; mm Hg), heart rate (beats per minute), temperature (°C) and consciousness (measured by Glasgow Coma Scale total score). We also included diastolic blood pressure, which is not part of NEWS2. Measurement time was chosen as the most completed set of parameters; measurements done 48 hours and 7 days prior to events were not included due to missing and inconsistent data. Heart rhythm was included 24 hours prior to event due to measurement recording by cardiac resuscitation team (CRT).

Outcomes

The primary outcome was patients' critical status following assessment on admission or 24 hours prior to a critical event. Outcomes were critical events categorised as in-hospital death, transfer to ICU, developing cardiac arrest or medical emergency. A medical emergency was defined as deterioration, excluding cardiac arrest, requiring a patient to be seen by a critical care outreach team due to vasovagal attack, breathing difficulty, bleeding, loss of consciousness, seizure, cardiac tamponade, chest pain or prearrest rhythm.

Data processing

Data were extracted for patients admitted from January to December 2020. Patients' demographics, physiological parameters, NEWS2 score, death and transfer to ICU were extracted from EHR. Diagnosis and comorbidities were gathered using the International Classification of Diseases 10th Revision coded data. Cardiac arrest and medical emergency were extracted from the CRT database. Patients with positive COVID-19 cases were identified from the COVID-19 pathology data, a daily updated database submitted to NHS England. Data for critical events and COVID-19 cases were linked to extracted EHR data using SQL by clinical data analysts in the hospital. Data were pseudonymised, transferred to the principal investigator (PI; BA) via the NHS network and then moved to University College London Data Safe Haven (DSH) for data analysis. The DSH is a secure database

Table 2 General characteristics of patients

Variables	Descriptive (n=6143)
Age (years)	63.7±14.47
Gender (%)	
Male	4239 (69)
Female	1904 (31)
Specialty (%)	
Cardiology	3817 (62)
Cardiothoracic surgery	2020 (33)
Congenital heart disease	184 (3)
Cardiac surgery	122 (2)
Patients with COVID-19 (%)	248 (4)
Cardiology and cardiothoracic surgery (%)	100 (40)
Other (oncology, haematology, respiratory and thoracic surgery) (%)	148 (60)
NEWS2 numerical parameters	
Systolic blood pressure (mm Hg)	128.3±23.2
Mean arterial pressure (mm Hg)	91.21±14.1
Pulse rate (beats per minute)	74.43±16.7
Temperature (°C)	36.5±0.6
Oxygen saturation (%)	96.7±2.23
Diastolic blood pressure (mm Hg)	72.7±11.9
NEWS2	1.5±1.7
Outcomes (%)	
In-hospital mortality	743 (12)
ICU admission	921 (15)
Cardiac arrest	117 (2)
Medical emergency	160 (3)

ICU, intensive care unit; NEWS2, National Early Warning Score 2.

Table 3 Comparison between categorical parameters of study population

Characteristic	NEWS2 categories Low (0–4) n (%)	Moderate (5–6) n (%)	High (≥7) n (%)	P value
Specialty				
Cardiology	3431 (90)	249 (7)	128 (3)	<0.001
Cardiothoracic surgery	1931 (83)	86 (4)	10 (0.5)	<0.001
Congenital heart disease	164 (89)	17 (9)	3 (2)	<0.001
Cardiac surgery	113 (92)	6 (5)	3 (2)	<0.001
Outcomes				
In-hospital mortality	597 (80)	62 (8)	84 (12)	<0.001
ICU admission	850 (92)	36 (4)	35 (4)	0.013
Cardiac arrest	85 (73)	14 (12)	18 (15)	<0.001
Medical emergency	140 (87)	12 (8)	8 (5)	<0.001

ICU, intensive care unit; NEWS2, National Early Warning Score 2.

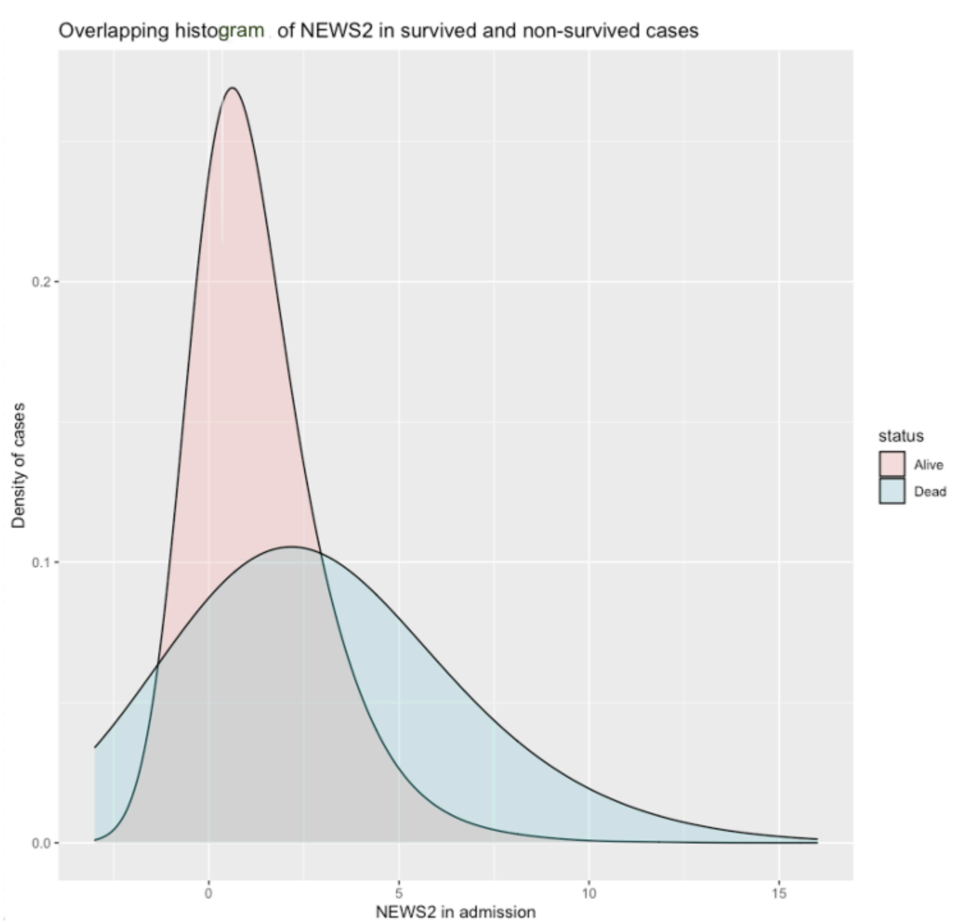


Figure 2. NEWS2 in survived and non-survived cases.

Figure 2 National Early Warning Score 2 (NEWS2) in survived and non-survived cases.

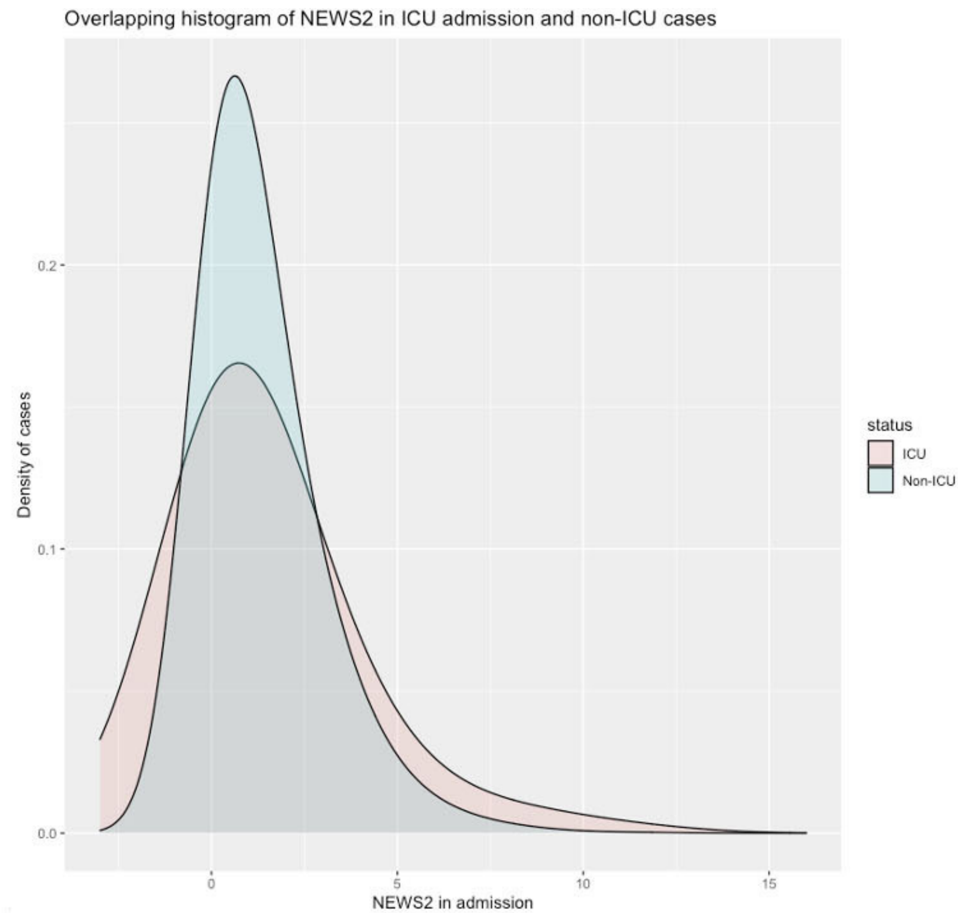


Figure 3 National Early Warning Score 2 (NEWS2) in intensive care unit (ICU) admission and non-ICU cases.

system with restricted access to the PI and research team via safe gateway technology.

Statistical analysis

Analysis was done in the DSH using the R program. Data cleaning and preprocessing were done by BA. Dependent and independent values, categorical and missing values and splitting data into training and test data sets were all addressed prior to analysis.

Statistical significance was defined as p value <0.5 using two-tailed tests.

The missing data in NEWS2 scores were assessed based on the reason for missing data. The categorical variables are presented as percentages (count), and the continuous variables are presented as the mean±SD. The normality analysis of the data was assessed using box plots for the frequencies. The intergroup difference between categorical variables was evaluated using Pearson's χ^2 test. The difference between groups was compared using the Mann-Whitney U test for non-normally distributed data.

The correlations between NEWS2 and physiological parameters and outcomes were evaluated using the Pearson correlation coefficient. Correlation coefficient values have a range between −1.0 and 1.0, where −1.0 shows perfect negative correlation and 1.0 indicates perfect positive correlation. To supplement the model with parameters that could improve the prediction, we split data into training and testing data sets using the train/test method (70% for training and 30% for testing) using decision tree through classification tree method by CART package in R. Univariate and multivariate logistic regression analyses were conducted to assess the association between score and outcomes. The predictive value of NEWS2 and supplemented model for hospital death, transfer to ICU, cardiac arrest and medical emergency were evaluated using the receiver operating characteristic (ROC) analysis. ROC is a graphical plot that shows the diagnostic ability of binary classifiers by plotting sensitivity against specificity for cut-off values used to guide the discrimination. The value of the areas under the ROC curve (AUC) was measured. The cut-off points of the models were assessed using Youden's index: sensitivity, specificity and positive and negative predictive values. The AUC values were interpreted using the reported criteria by Fischer *et al*: AUC >0.9, 0.7–0.9 and 0.5–0.6 indicate high, moderate and low predictive accuracy, respectively.[17]

Patient and public involvement

Patients were not involved in setting the research question, outcome measures nor in the design of the study. Patients were not involved in interpretation, writing up of the results. There is no plan for the results to be disseminated to the patient population affected.

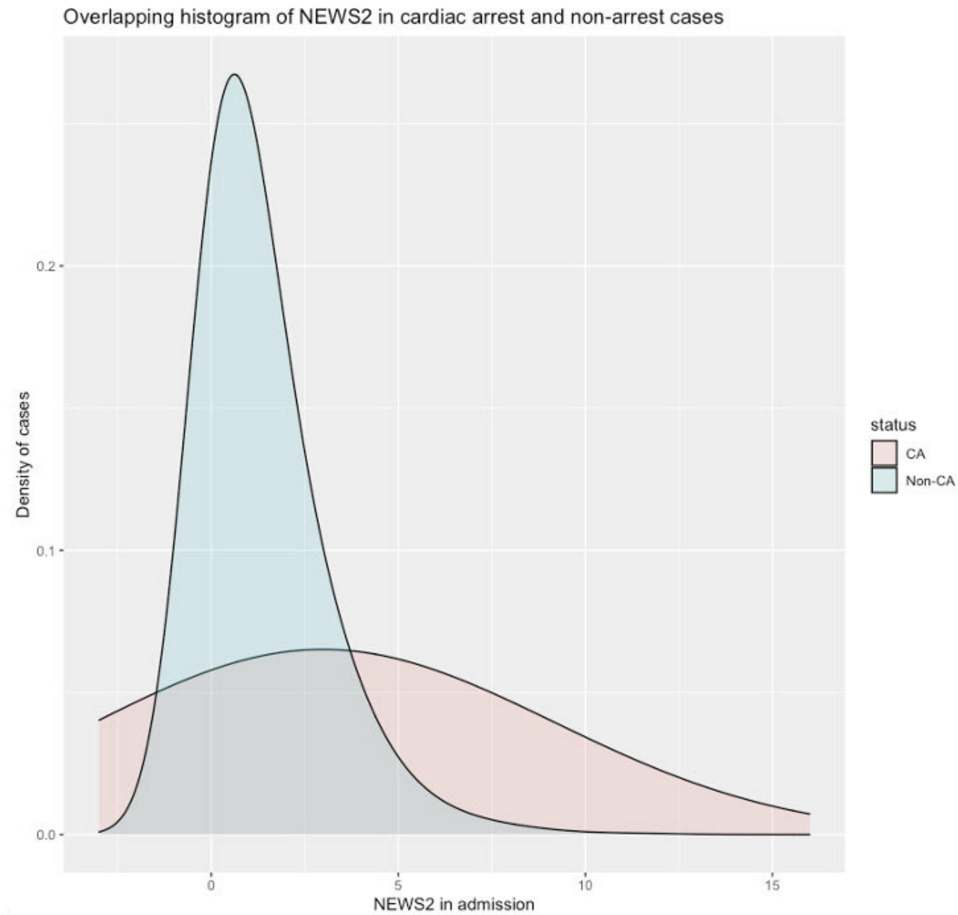


Figure 4 National Early Warning Score 2 (NEWS2) in cardiac arrest (CA) and non-arrest cases.

RESULTS

Baseline characteristics

The initial cohort comprised 16 978 admitted patients, forming 40 901 encounters and 68 867 admissions to the wards from various specialties in oncology, cardiology, medicine and surgery. Patients with a primary CVD diagnosis were 7313 (36%), 14 798 encounters and 24 792 ward admissions. Patients with missing NEWS2 or physiological parameter values were 21% of the total cohort and 16% of cardiac patients. Missing NEWS2 scores were unrelated to specific values in the data set and were more likely to result from failure to record or loss in transit. In this case, they were regarded as missing completely as random[17] and were removed.[18] Included cardiac patients were 6143 patients admitted under cardiology, cardiothoracic surgery, congenital heart diseases or cardiac surgery specialties (online supplemental appendix 1). Patients with COVID-19 were 248 (4%), 40% were cardiac patients.

The mean age of the cardiac population is 63.73+14.47, and 69% were males. The in-hospital mortality was 12% (743 patients), ICU admission was 15% (921 patients), and 117 cardiac arrests and 160 medical emergencies. The characteristics of the study population are tabulated in table 2. The difference between dead cases, patients admitted to ICU, who developed cardiac arrest, or medical emergencies, and those who did not develop critical outcomes according to the NEWS2 scoring category were statistically significant ($p<0.001$). The comparison is tabulated in table 3.

The mean of NEWS2 in death cases was higher by a small difference than alive patients (difference=1.005, $p<0.001$). Between ICU admission and non-admitted cases, the mean was similar (difference=0.01, $p<0.09$). Between cardiac arrests and non-arrest cases, and medical emergency and stable cases, there was a small variation (difference=1.99, $p<0.001$ and difference=0.99, $p<0.001$, respectively) (figures 2–5).

Using the correlation matrix between parameters measured on admission and outcomes, we found a positive correlation between temperature with heart rate (0.32); respiration rate, heart rate and death with NEWS2 (0.41, 0.31, 0.30); and death with cardiac arrest (0.31). In the parameters 24 hours prior to critical events, there was a strong correlation of SpO_2 with systolic pressure, CVPU (new Confusion, Voice, Pain, Unresponsive) with NEWS2 (0.42, 0.41, 0.42), and systolic pressure, SpO_2 and death with age (0.30, 0.31, 0.34). Cardiac rhythm is strongly correlated with cardiac arrest and death (0.51, 0.90) (figure 6 and online supplemental appendix 2).

Regarding the discrimination, NEWS2 on admission showed moderate to low predictive accuracy with death, ICU admission, cardiac arrest and medical emergency (AUC: 0.63, 0.56, 0.69 and 0.63, respectively), while NEWS2 24 hours before event showed low predictive value (AUC: 0.58, 0.63, 0.54 and 0.56, respectively). In patients with COVID-19,

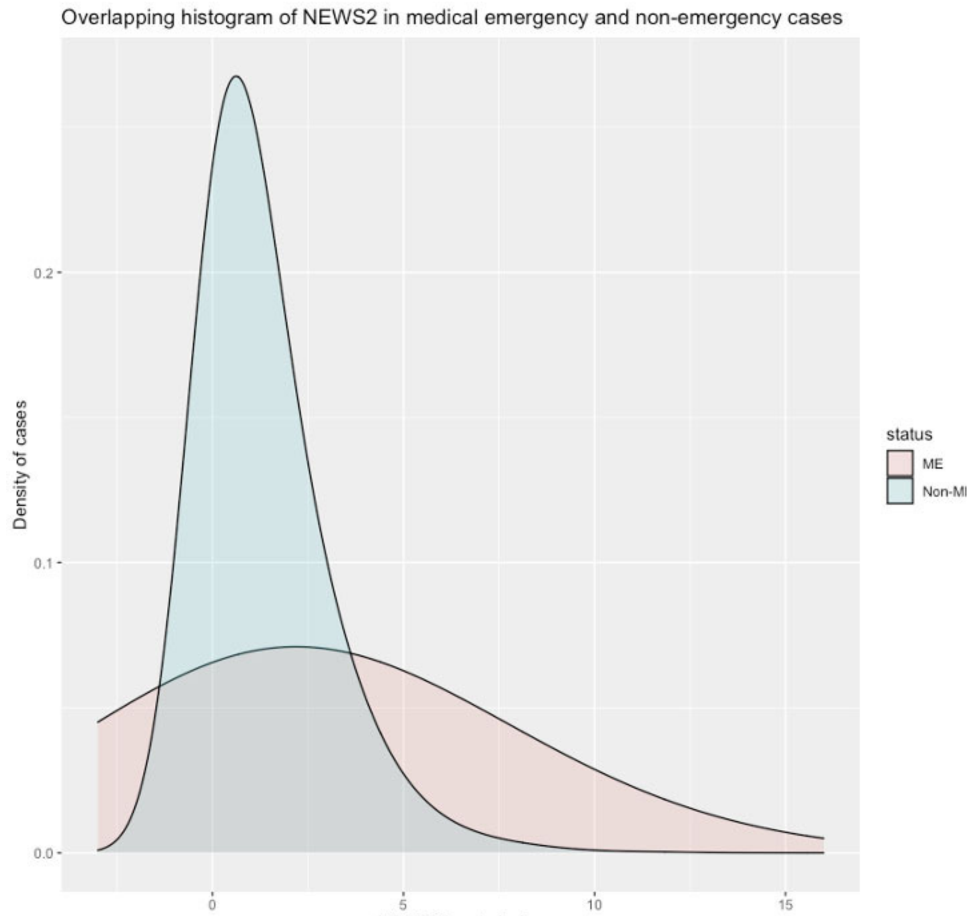


Figure 5 National Early Warning Score 2 (NEWS2) in medical emergency (ME) and non-emergency cases.

NEWS2 on admission showed good to poor performance (AUC: 0.64, 0.51, 0.81 and 0.80). When NEWS2 was supplemented with age, there was insignificant change in the predictive performance for all patients (AUC: 0.63, 0.53, 0.73 and 0.64) However, there was significant improvement for patients with COVID-19 (AUC: 0.96, 0.69, 0.87 and 0.88). This was also true for the model of NEWS2 supplemented with heart rhythm for the cardiac patients (AUC: 0.75, 0.84, 0.95 and 0.94). The calculated optimum cut-off value for NEWS2 was >5 which showed sensitivity for NEWS2 model of 20% and specificity of 94%, while for the model supplemented with heart rhythm sensitivity was 30% and specificity was 85% (table 4 and online supplemental appendix 3).

DISCUSSION

Our retrospective study is among the first to evaluate the prognostic power of the digital EWS (NEWS2) in a patient with CVD in a specialist cardiac setting. We were able to validate NEWS2 in patients with COVID-19 in the cardiac setting. The main findings of the study reveal that: (1) NEWS2 is inadequate on its own to predict deterioration in patients with CVD in the examined specialist cardiac setting; and (2) adjustment of the tool by supplementing with positively correlated parameters can strengthen the prognostic performance and therefore reduce the burden of critical events associated with CVD.

Our findings were consistent with a previous study in patients with chest pain[19] while contrary to findings by a study that examined a subset of patients with CVD in a single hospital.[15] MEWS (modified early warning score) showed low predictive accuracy in patients with chest pain in the ED.[19] However, good discrimination was found of ViEWS in a subset of patients in Canada a decade ago and using a distinct EWS than the current NEWS2.[15] Studies were limited in number, scope and population and varied in EWS models. Patients with 'normal' vital signs may be sicker than they look through traditional routine monitoring.[20] When EWS was explicitly developed for postcardiac surgery patients, prognostic accuracy was excellent in predicting ICU mortality.[16] It included a range of organ system-specific parameters that correlate with cardiac surgery outcomes, such as lactic acid, FiO_2 and platelets. Relevant parameters were found using machine learning, including clinical signs and heart rate variability, to improve scoring systems for adverse cardiac events.[19] The cardiac rhythm combined with NEWS2 in our study, and the heart rate variability, such as the average of the instantaneous heart rate or ratio of low frequency (LF) power to high frequency (HF) power (LF/HF) selected by Liu

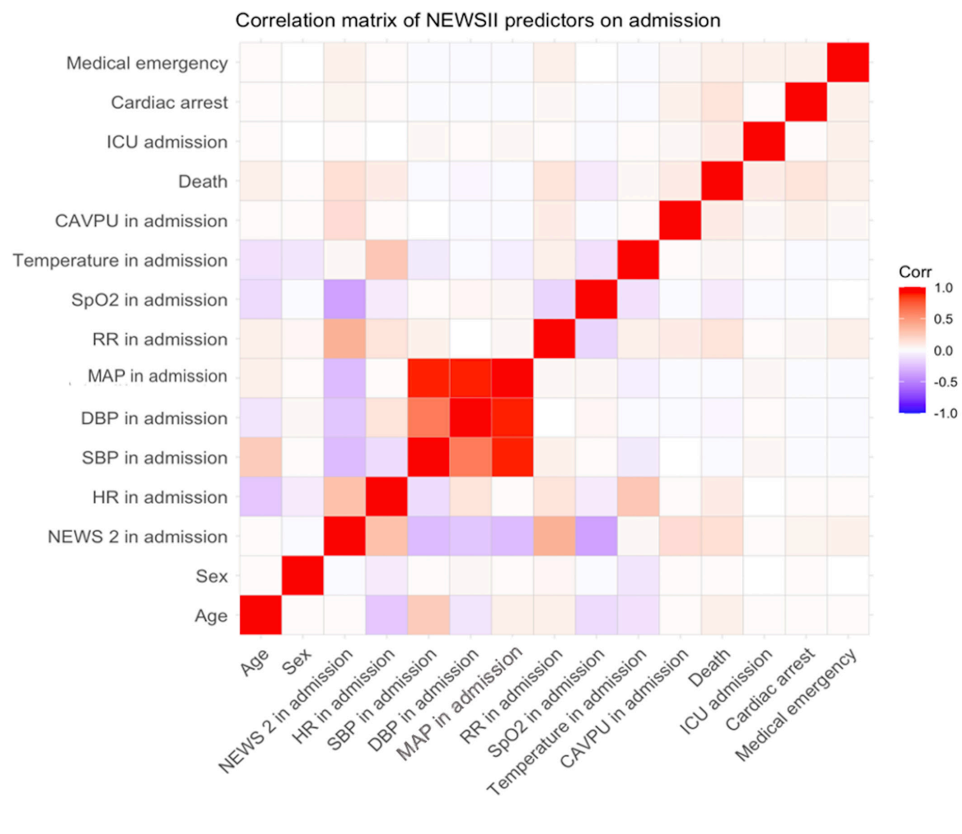


Figure 6 Correlation matrix using Pearson's correlation coefficient between parameters and National Early Warning Score 2 (NEWS2) on admission and outcomes. DBP, diastolic blood pressure; ICU, intensive care unit; SBP,systolic blood pressure; CVPU, new Confusion, Voice, Pain, Unresponsive; RR, Respiration rate; MAP, Mean arterial blood pressure.

et al,[19] may not be routinely measured or readily available for clinicians as SBP or temperature. Yet, their predictive value indicates the need for highly illness-correlated parameters to be present for clinicians through facilitating timely and thorough assessment.

The subset of admitted patients with COVID-19 showed improvement in NEWS2 when the model was adjusted with age. The finding is supported by a study that reported low to moderate discrimination of NEWS2 for the severity of COVID-19 disease.[21] Adjusting the model with age alone did not improve prediction in the study by Carr *et al*, as shown in our findings.[9] However, supplementing with age, routinely collected blood and physiological parameters enhanced discrimination in a multisite (UK and non-UK hospitals) study,[9] which seems to further support our results yet may detect the need for additional criteria adjustment. Our study indicates that EWS are not a stand-alone rapid assessment tool as it accords with reported limitations in rapid risk scores for predicting cardiovascular complications.[22 23] In clinical settings, and before implementing EWS, clinicians look for signs of abnormality related to body organs affected, disease

Table 4 NEWS2 as a predictor of critical events compared with supplemented NEWS2

Model	Death AUC (95% CI)	ICU admission AUC (95% CI)	Cardiac arrest AUC (95% CI)	Medical emergency AUC (95% CI)
Cardiac patients				
NEWS2 (admission)	0.63 (0.58–0.67)	0.56 (0.51–0.62)	0.69 (0.65–0.74)	0.63 (0.57–0.67)
NEWS2 (24 hours before outcome)	0.58 (0.53–0.65)	0.63 (0.58–69)	0.54 (0.51–0.67)	0.56 (0.52–0.60)
NEWS2+age	0.63 (0.58–0.69)	0.53 (0.50–0.59)	0.73 (0.68–0.79)	0.64 (0.59–0.68)
NEWS2+rhythm	0.75 (0.67–0.80)	0.84 (0.78–0.88)	0.95 (0.89–0.98)	0.94 (0.89–0.97)
Patients with COVID-19				
NEWS2 (admission)	0.64 (0.59–0.69)	0.51 (0.50–0.56)	0.81 (0.75–0.86)	0.80 (0.74–0.86)
NEWS2+age	0.96 (0.89–0.98)	0.69 (0.63–0.76)	0.87 (0.81–0.94)	0.88 (0.82–0.94)

AUC, area under the receiver operating characteristic curve; ICU, intensive care unit; NEWS2, National Early Warning Score 2.

pathophysiology or procedure side effect to critically assess the situation. Systems that existed to stratify the risk of long-term cardiac complications have been successfully validated and used for years, such as the thrombolysis in myocardial infarction score[24] and GRACE.[25] They included cardiac disease variables: heart rate variabilities and serum cardiac biomarkers, which may not be available routinely or at the first admission presentation for rapid assessment. It was also observed that combining nurses' objective assessment with traditional EWS in the Dutch-Early-Nurse-Worry-Indicator-Score improved the prediction of ICU admission and mortality in surgical patients.[26] Therefore, thorough tracking of short-term deterioration parameters to develop decisive intelligent scoring systems will potentially outperform EWS in various diseases and settings.

It is essential to consider possible issues in developing EWS for specialised subgroups. A potential complexity may be present when having a variety of multiple parameters measured at various times during admission to form a scoring system that is meant to be simplistic and standardised. In addition, the endpoints favoured by researchers in validation studies may not be the ideal points to measure triggered EWS against. In the clinical application of EWS, a high score triggers an action to prevent a critical event.[27] In the event of clinically intervening at the right time, the examined adverse events may not occur. Prior to reaching or while preventing an adverse event, precise and proper deterioration endpoints may be more fitted than traditionally studied outcomes, such as death and ICU admission. At the current time of available EHR integration and data science techniques, it is possible and may be more valid to identify and define appropriate critical illness endpoints to examine EWS against.

We assessed the performance of digitally integrated EWS; the integration generates NEWS2 in patients' charts from remotely captured parameters by automated monitoring. EHR integration and automation can improve the accuracy and alerting promptness.[28] We were able to extract a good sample of patients with CVD and identify patients diagnosed with COVID-19 despite the missingness of some NEWS2 recordings. The issue in recording completion could be due to a lack of staff adherence to routine and timely monitoring, as each measurement would be automatically transmitted to patients' charts.

Therefore, careful and selective modelling of algorithms from parameters that can be available routinely and reflect significant clinical meaning is needed. The validity of NEWS2 in specialist settings like cardiology indicates the need for either score enhancement or systemic supplementing and finer endpoints definition for better detection. Studying the clinical environment from a practical side of EWS will explain the adoption and implementation role in the success or failure of EWS. From various specialties, clinicians' involvement in models' development and validation is invaluable to produce a higher accuracy and finely clinical expertise-born warning scores.

CONCLUSION

The EWS (NEWS2) in patients with CVDs is suboptimal to predict deterioration early. Adjusting EWS with variables that strongly correlate with critical cardiovascular outcomes will improve the early scoring models. Thorough tracking of parameters in EHRs and data availability can support the generation of decisive, intelligent models for a readily feasible system in routine clinical work. There is a need for defining and revising critical endpoints and the involvement of clinicians in models' development that reflect a significant meaning for deterioration detection. Further validation and implementation studies in cardiac specialist settings and other specialist subgroups are required to investigate methods needed to enhance the capacity of EWS where it is least investigated.

Twitter Amitava Banerjee @amibanerjee1

Contributors AB, DM and TB conceived the study. BA carried out the data collection with guidance from DM. BA conducted the analysis and interpretation of findings with revisions from AB. BA wrote the manuscript. AB, DM, TB and RP contributed to the interpretation and revision of the manuscript. AB was guarantor for this work.

Funding BA has received PhD funding from the Saudi Arabian Cultural Bureau to conduct the study (grant number: ELP003964).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study is registered and approved by the Health Research Authority (HRA) and Health and Care Research Wales (HCRW). Approval was obtained from Stanmore Research Ethics Committee in London (REC reference: 20/PR/0286). A confirmation of capacity has been received from St Bartholomew's Hospital. Data collected are essential for patient care and were sufficiently anonymised; therefore, it was not considered necessary to obtain consent from patients.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Baneen Alhmoud http://orcid.org/0000-0002-7165-3717
Riyaz Patel http://orcid.org/0000-0003-4603-2393
Amitava Banerjee http://orcid.org/0000-0001-8741-3411

REFERENCES

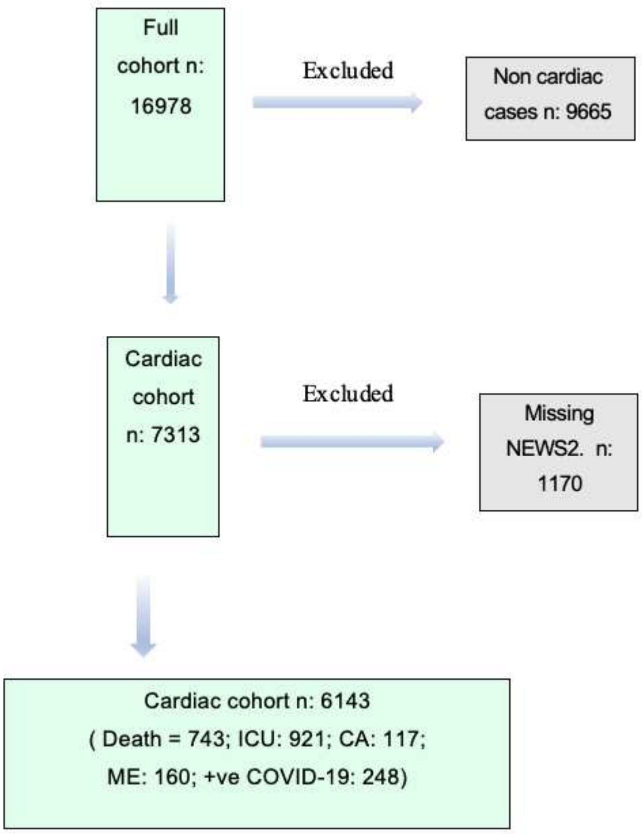
1 Mozaffarian D. Global scourge of cardiovascular disease. *J Am Coll Cardiol* 2017;70:26–8.

2 Inada-Kim M, Nsutebu E. News 2: an opportunity to standardise the management of deterioration and sepsis. *BMJ* 2018;360:k1260.
3 BHF. n.d. UK factsheet. british heart foundation. 2019;(April):1–21.
4 Herlitz J, Bång A, Ekström L, *et al*. A comparison between patients suffering in-hospital and out-of-hospital cardiac arrest in terms of treatment and outcome. *J Intern Med* 2000;248:53–60.
5 Sandroni C, Nolan J, Cavallaro F, *et al*. In-Hospital cardiac arrest: incidence, prognosis and possible measures to improve survival. *Intensive Care Med* 2007;33:237–45.
6 Aragam KG, Tamhane UU, Kline-Rogers E, *et al*. Does simplicity compromise accuracy in ACS risk prediction? A retrospective analysis of the TIMI and grace risk scores. *PLoS One* 2009;4:e7947.
7 Shariff N, Aleem A, Singh M, *et al*. AF and venous thromboembolism - pathophysiology, risk assessment and CHADS-vasc score. *J Atr Fibrillation* 2012;5:649.
8 Smith GB, Prytherch DR, Meredith P. 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;84:465–70.
9 Carr E, Bendayan R, Bean D, *et al*. Supplementing the national early warning score (NEWS2) for anticipating early deterioration among patients with COVID-19 infection. *MedRxiv [Internet]* 2020.
10 Day T, Oxton J. The National early warning score in practice: a reflection. *Br J Nurs* 2014;23:1036–40.
11 Royal College of Physicians of London. *National early warning score (NEWS): standardising the assessment of acute-illness severity in the NHS*. Royal College of Physician, 2012.
12 Dundar ZD, Kocak S, Girisgin AS. Lactate and NEWS-L are fair predictors of mortality in critically ill geriatric emergency department patients. *Am J Emerg Med* 2020;38:217–21.
13 Alhmoud B, Bonnici T, Patel R, *et al*. Performance of universal early warning scores in different patient subgroups and clinical settings: a systematic review. *BMJ Open* 2021;11:e045849.
14 Eckart A, Hauser SI, Kutz A, *et al*. Combination of the national early warning score (news) and inflammatory biomarkers for early risk stratification in emergency department patients: results of a multinational, observational study. *BMJ Open* 2019;9:e024636.
15 Kellett J, Kim A. Validation of an abbreviated vitalpacvalidation of an abbreviated vitalpac. *Resuscitation* 2012;83:297–302.
16 Badreldin AMA, Doerr F, Bender EM, *et al*. Rapid clinical evaluation: an early warning cardiac surgical scoring system for hand-held digital devices. *Eur J Cardiothorac Surg* 2013;44:992–7;
17 Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 2003;29:1043–51.
18 Katuwal S, Knadel M, Norgaard T, *et al*. Predicting the dry bulk density of soils across denmark: comparison of single-parameter, multi-parameter, and vis–NIR based models. *Geoderma* 2020;361:114080.
19 Liu N, Koh ZX, Goh J, *et al*. Prediction of adverse cardiac events in emergency department patients with chest pain using machine learning for variable selection. *BMC Med Inform Decis Mak* 2014;14:75.
20 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.
21 Kostakis I, Smith GB, Prytherch D, *et al*. The performance of the national early warning score and national early warning score 2 in hospitalised patients infected by the severe acute respiratory syndrome coronavirus 2 (sars-cov-2). *Resuscitation* 2021;159:150–7.
22 Hollander JE, Robey JL, Chase MR, *et al*. Relationship between a clear-cut alternative noncardiac diagnosis and 30-day outcome in emergency department patients with chest pain. *Acad Emerg Med* 2007;14:210–5.
23 Manini AF, Dannemann N, Brown DF, *et al*. Limitations of risk score models in patients with acute chest pain. *Am J Emerg Med* 2009;27:43–8.
24 Antman EM, Cohen M, Bernink PJ, *et al*. The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making. *JAMA* 2000;284:835–42.
25 Eagle KA, Lim MJ, Dabbous OH, *et al*. A validated prediction model for all forms of acute coronary syndrome: estimating the risk of 6-month postdischarge death in an international registry. *JAMA* 2004;291:2727–33.
26 Douw G, Huisman-de Waal G, van Zanten ARH, *et al*. Nurses' "worry" as predictor of deteriorating surgical ward patients: a prospective cohort study of the dutch-early-nurse-worry-indicator-score. *Int J Nurs Stud* 2016;59:134–40.
27 Pedrosa I, Browning T, Kwon JK, *et al*. Response to COVID-19: minimizing risks, addressing challenges and maintaining operations in a complex academic radiology department. *J Comput Assist Tomogr* 2020;44:479–84.
28 National Institute For Health And Care Excellence (NICE). National early warning score systems that alert to deteriorating adult patients in hospital. *Medtech Innovation Briefing* 2020:1–18.:.
29 Royal College of Physicians of London. *NHS england approves use of national early warning score (NEWS) 2 to improve detection of acutely ill patients*. Royal College of Physician, 2017.
30 Schreuder HT, Czaplewski R, Bailey RG. Combining mapped and statistical data in forest ecological inventory and monitoring – supplementing an existing system. *Environ Monit Assess* 1999;56:269–91. 10.1023/A:1005984426987 Available: https://doi.org/10.1023/A:1005984426987
31 Altman DG, Royston P. What do we mean by validating a prognostic model? *Statist Med* 2000;19:453–73.
32 Lockwood JM, Thomas J, Martin S, *et al*. AutoPEWS: automating pediatric early warning score calculation improves accuracy without sacrificing predictive ability. *Pediatr Qual Saf* 2020;5:e274.
33 Khanna AK, Hoppe P, Saugel B. Automated continuous noninvasive ward monitoring: future directions and challenges. *Crit Care* 2019;23:194.

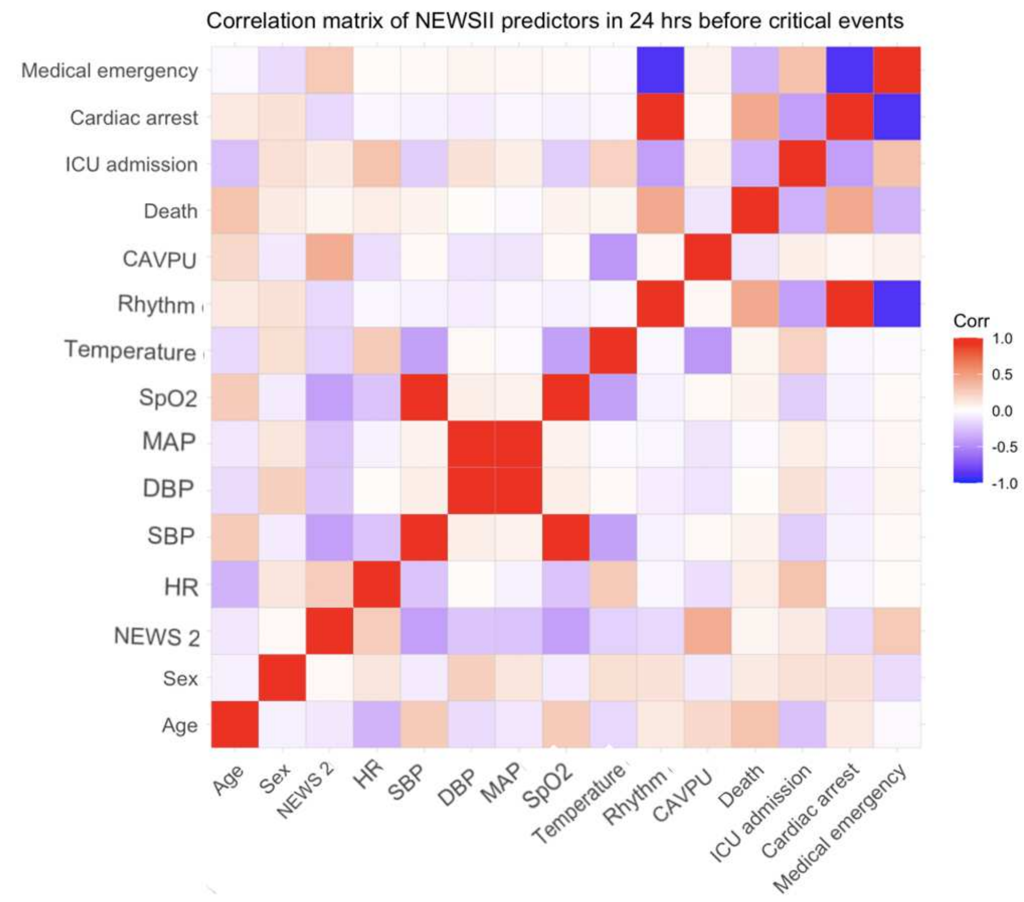
Performance of digital Early Warning Score (NEWS2) in a cardiac specialist setting: retrospective cohort study.

Appendices

1- Flowchart of patients' cohort and data sources



Abbreviations: n: number of patients; ICU: cases admitted to Intensive care unit; ME: medical emergency cases; CA: cardiac arrest cases; |+veCOVID-19: cases diagnosed with COVID-19.

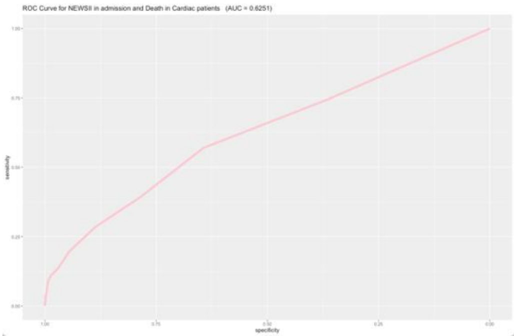


2. Correlation matrix using Pearson's correlation coefficient for parameters and NEWS2 24 hrs before outcomes.

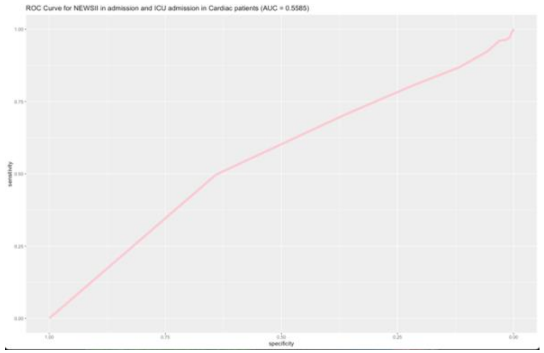
3- Predictive ability of NEWS2.

A. Predictive ability of NEWS2 on admission for death, ICU admission, cardiac arrest, and medical emergency.

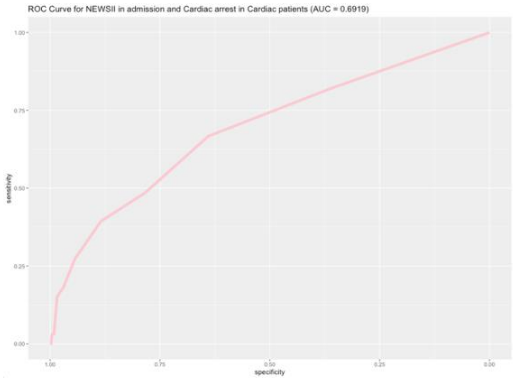
1- Death AUC:0.63



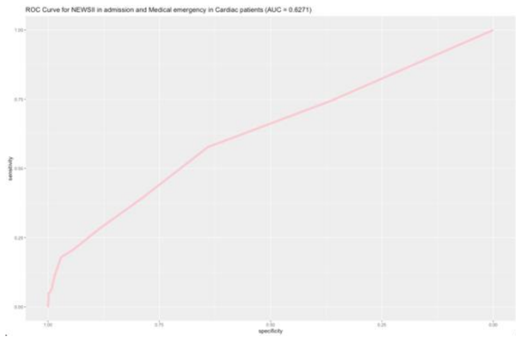
2. ICU. AUC:0.56



3. CA AUC:0.69

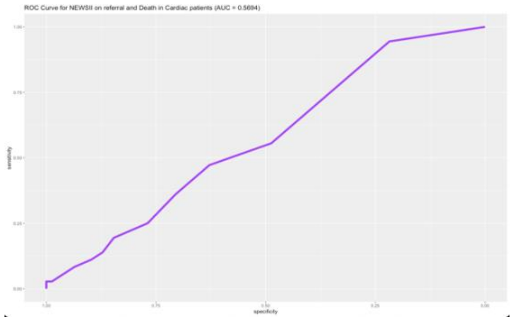


4. ME AUC: 0.63

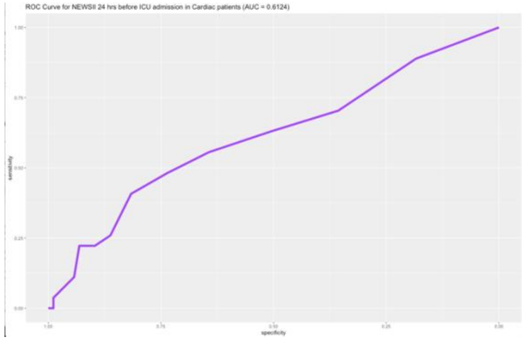


B. NEWS2 predictive ability 24 hours before critical events

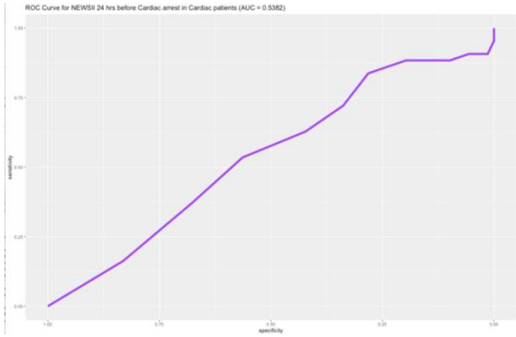
1. Death AUC:0.57



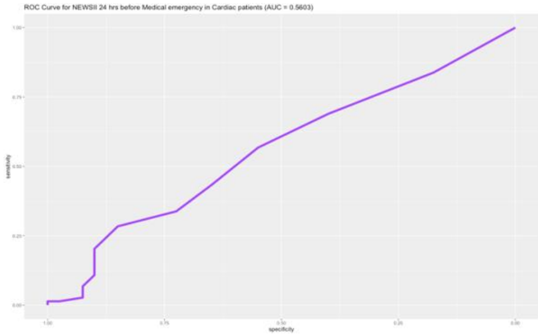
2. ICU. AUC:0.61



3. CA. AUC:0.53

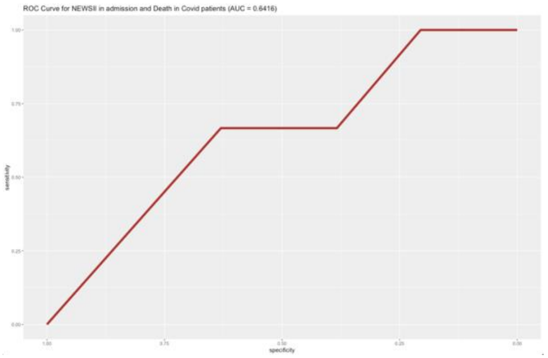


4. ME. AUC:0.56

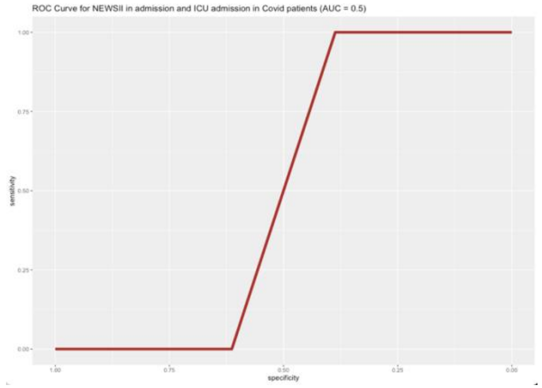


C. NEWS2 predictive ability on admission in patients with COVID-19

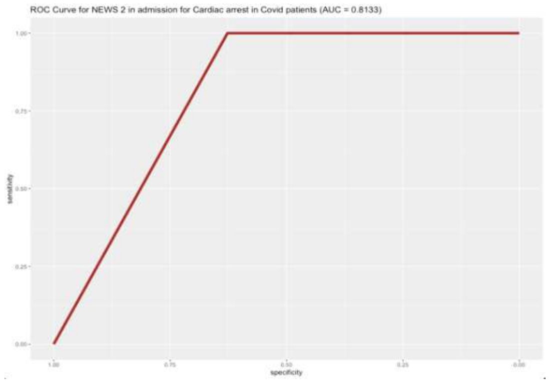
1- Death AUC:0.64



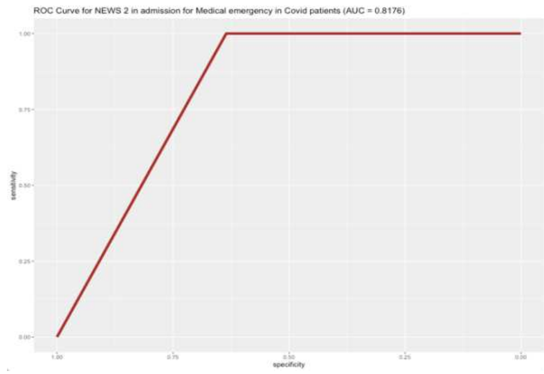
2. ICU AUC:0.5



3.CA AUC:0.81

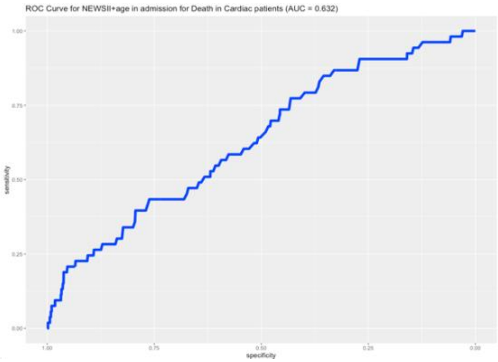


4. ME. AUC:0.81

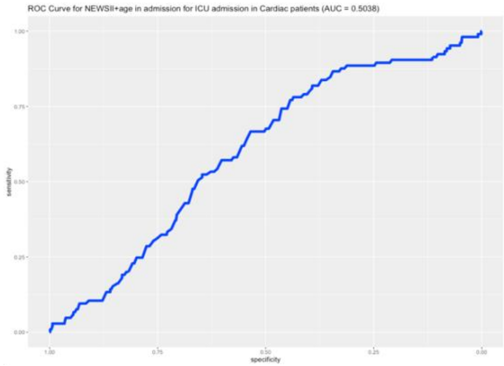


F. NEWS2+age predictive ability in cardiac patients

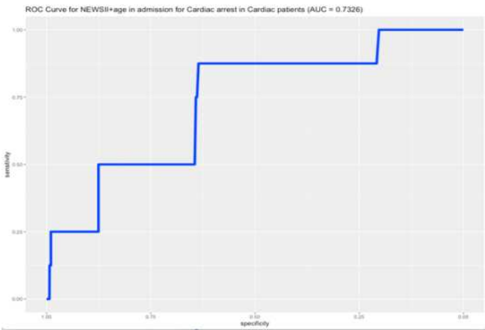
1- Death AUC: 0.63



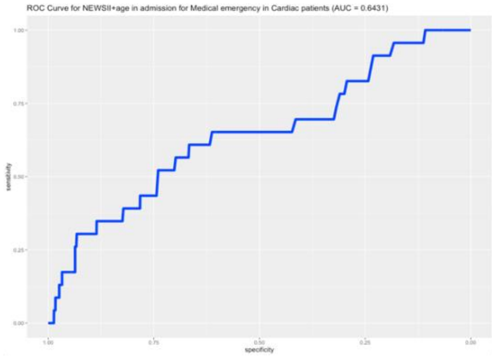
2. ICU AUC:0.5



3.CA AUC:0.73

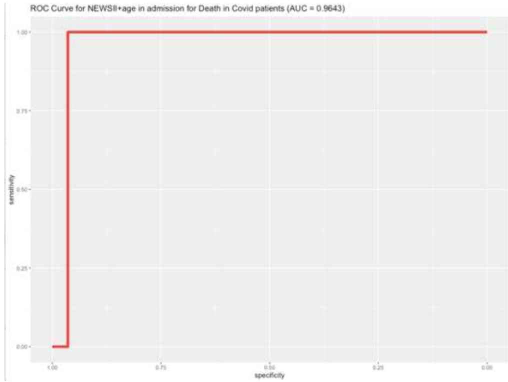


4.ME. AUC:0.64

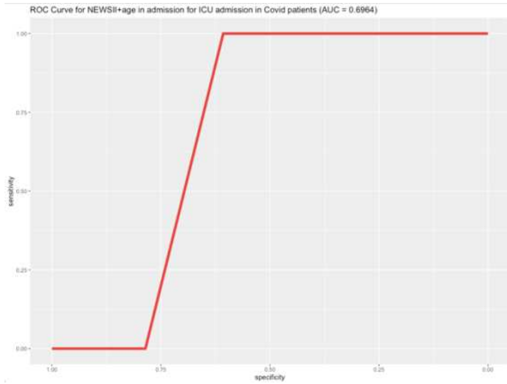


G. NEWS2+ age predictive ability in COVID-19 patients

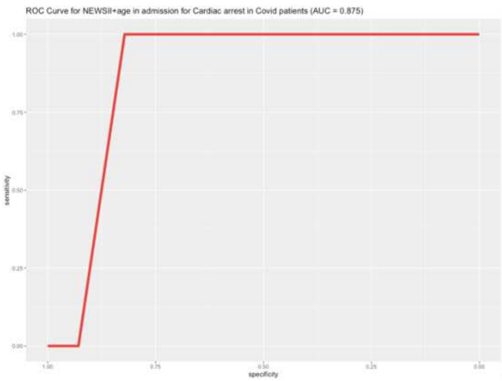
1.Death AUC:0.96



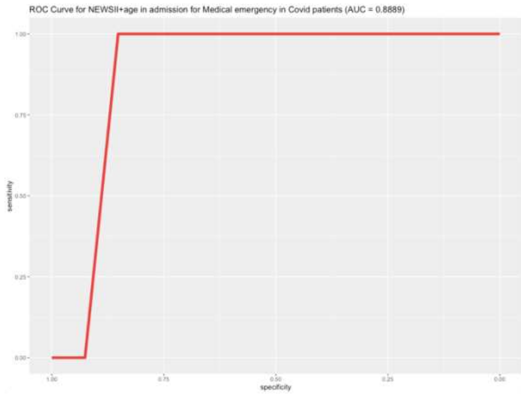
2. ICU AUC: 0.7



3.CA AUC:0.87

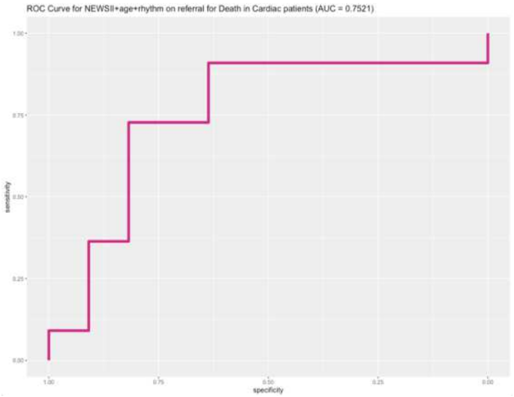


4.ME AUC:0.88

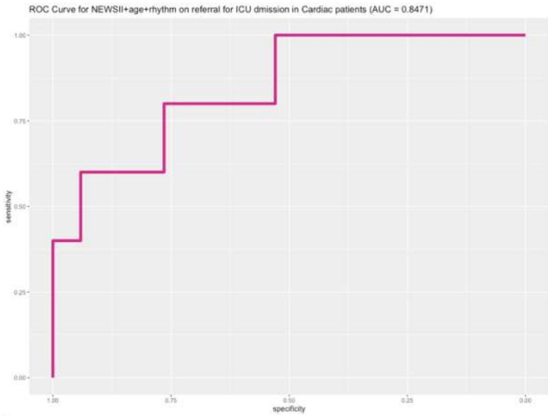


H. NEWS2 +age + cardiac rhythm predictive ability in cardiac patients

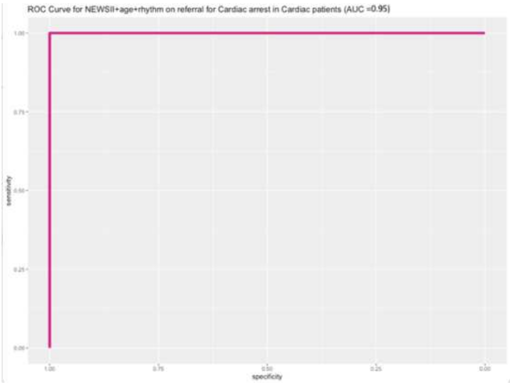
1- Death AUC: 0.75



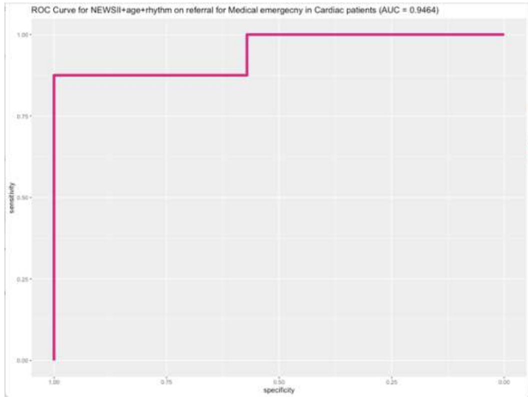
2. ICU. AUC:0.84



3.CA AUC: 0.95



4.ME AUC:0.94



Abbreviations: AUC: are under receiving curve, ICU: admission to intensive care unit, CA: cardiac arrest and ME: medical emergency.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	3
	4	Study objectives and hypotheses	3 and 4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	4
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4
	9	Whether participants formed a consecutive, random or convenience series	4
Test methods	**10a**	Index test, in sufficient detail to allow replication	4 and5
	10b	Reference standard, in sufficient detail to allow replication	4 and 5
	11	Rationale for choosing the reference standard (if alternatives exist)	5
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	5 - 7
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	5 - 7
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	5-7 and appendix 2
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	5-7 and appendix 2
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	5
	15	How indeterminate index test or reference standard results were handled	5
	16	How missing data on the index test and reference standard were handled	5 and 6
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	5
	18	Intended sample size and how it was determined	4 and 5
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Appendix 1
	20	Baseline demographic and clinical characteristics of participants	5 and 6
	21a	Distribution of severity of disease in those with the target condition	6
	21b	Distribution of alternative diagnoses in those without the target condition	
	22	Time interval and any clinical interventions between index test and reference standard	7 and 8
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	7 and 8
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	7
	25	Any adverse events from performing the index test or the reference standard	
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	8 and 9
	27	Implications for practice, including the intended use and clinical role of the index test	9 and 10
OTHER INFORMATION			10
	28	Registration number and name of registry	
	29	Where the full study protocol can be accessed	
	30	Sources of funding and other support; role of funders	10

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on [http://www.equator-network.org/reporting-guidelines/stard.](http://www.equator-network.org/reporting-guidelines/stard)