Risk stratification in emergency medicine: towards improved age and sex adjusted risk assessment

Candel, B.G.J.

PART

II

DEVELOPMENT AND EXTERNAL VALIDATION OF AN AGE AND SEX ADJUSTED EARLY WARNING SCORE


EMERGENCY

7

THE IMPACT OF AGE ON PREDICTION PERFORMANCE OF NATIONAL EARLY WARNING SCORE ON ARRIVAL AT EMERGENCY DEPARTMENTS – DEVELOPMENT AND EXTERNAL VALIDATION

Søren Kabell Nissen
Bart GJ Candel
Christian H. Nickel
Evert de Jonge
Jesper Ryg
Søren Bie Bogh
Bas de Groot
Mikkel Brabrand

Published in the Annals of Emergency Medicine – 2022, Volume 79, Issue 4, p354-363.

ABSTRACT

Objective To investigate how age affects the predictive performance of the National Early Warning Score (NEWS) at arrival to the Emergency Department (ED) regarding in-hospital mortality and intensive care admission.

Methods International multicenter retrospective cohorts from two Danish and three Dutch ED. Development cohort: 14,809 Danish patients ≥ 18 years with at least systolic blood pressure or pulse measured from the Danish Multicenter Cohort (DMC). External validation cohort: 50,448 Dutch patients ≥ 18 years with all vital signs measured from the Netherlands Emergency Department Evaluation Database (NEED). Multivariable logistic regression was used for model building. Performance was evaluated overall and within age-categories: 18-64, 65-80 and >80 years.

Results In the DMC, a total of 2.5% died in-hospital and 2.8% were admitted to the ICU, compared to 2.8% and 1.6%, respectively, in the NEED. Age did not add information for prediction of intensive care admission but was the strongest predictor for in-hospital mortality. For NEWS alone, severe underestimation of risk was observed for patients >80 years, while overall Area Under Receiver Operating Characteristic (AUROC) was 0.82 (CI: 0.80-0.84) in the DMC and 0.75 (CI: 0.75-0.77) in the NEED. When combining NEWS with age, underestimation of risks was eliminated for patients >80 years and overall AUROC increased significantly to 0.86 (CI: 0.85-0.88) in the DMC and 0.82 (CI: 0.81-0.83) in the NEED.

Conclusions Combining NEWS with age improved the prediction performance regarding in-hospital mortality, mostly for persons aged above 80 and can potentially improve decision policies at arrival to EDs.

INTRODUCTION

Background

Early Warning Scores (EWS) aggregate vital signs to assess illness acuity and identify persons at risk of deterioration in the Emergency Department (ED).[130] Among EWS, the National Early Warning Score (NEWS) in particular has seen widespread implementation.[131] NEWS was developed after thorough review of 33 other EWS and combines six vital signs to a sum-score, developed for best possible discrimination for 24-hour mortality with an Area Under the curve of the Receiver Operating Characteristic (AUROC) of 0.89.[132] However, the ability to predict the prognosis of older ED patients using NEWS may be inadequate.[27, 42] One explanation may be that the risk of mortality associated with increasing vital sign abnormalities is larger in older persons,[133] or that vital signs discriminate worse for older compared to younger persons for a combined outcome of in-hospital mortality or ICU admission.[134] In development of NEWS, age was not included as a predictor because the impact of age on predictive performance was not considered convincing by developers.[132]

Importance

Older people in the ED could be at risk of delayed or incorrect treatment or wrong disposition, which may result in adverse outcomes. Previous studies in the ED settings either dichotomized or categorized age when investigating how to combine age with the similar Modified EWS.[133, 135] However, chronological age may not always reflect physiological age and risk may increase in a non-linear fashion. Thus, the gain or harm from combining NEWS with age remains uncertain.

7

Goals of This Investigation

The first objective of this study was to combine NEWS with age in a development cohort and determine predictive performance within age categories. The second objective was to externally validate the predictive performance.

METHODS

Study Design and Setting

This was an international multicenter cohort study based on existing cohorts from Emergency Departments. Analysis of predictive performance was performed for pre-planned subgroups, stratified by age into three categories (18-64, 65-80, and above 80 years).

Data collection conformed to the principles outlined in the Declaration of Helsinki. In Denmark, the study was approved by the Danish Data Protection Agency (file no. 2008-58-0035) and Danish Authorities for Health and Medicines (file no. 3-3013-1281/ 1). Under Danish law, retrospective register studies are exempt from the need for approval by an ethics committee. In the Netherlands, the study was approved by the medical ethics committee of the Máxima MC (file no. N21.007).

This study is presented in adherence to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) reporting standards.[136]

Selection of Participants: Development Cohort

We used data from the Danish multicenter cohort (DMC) and combined NEWS with age. The DMC includes patients ≥ 18 years with a Danish civil registration number. Patients were excluded if neither systolic blood pressure nor pulse were recorded as these were on average missing all 7 vital signs used in NEWS. Patients were consecutively sampled in relation to previous prospective studies at two level II emergency centers: University Hospital of Southwest Jutland: (2 October 2008 - 12 February 2009; 23 February 2010 - 26 May 2010; 1 June 2012 - 1 November 2011; 24 April 2013 - 9 December 2013), and Lillebaelt Hospital (1 January 2010 - 30 June 2010).[118, 137-139]

Selection of Participants: External validation cohort

For external validation, we used data from the Netherlands Emergency Department Evaluation Database (NEED, for more information see www.stichting-need.nl), the national quality registration database in the Netherlands. The NEED includes all patients ≥ 18 years, unless they objected to participation. Patients were only included if a complete set of vital sign measurements were present. For this study, data was available from a level III emergency center (1 January 2017 – 8 June 2019), and two emergency centers at level II (1 January 2019 – 12 January 2020; 1 January 2017 – 31 December 2019).

Measurements

In the DMC, systolic blood pressure, pulse, respiratory rate, temperature, oxygen saturation, use of supplementary oxygen, consciousness and basic demographic information were recorded by a nurse in structured form according to standard clinical practice for triage upon arrival. In the NEED, vital signs were recorded in a similar manner, and both vital sign measurement and basic demographic data were extracted from the Minimal

Data Set as described previously.[91] Each vital sign was then assigned a value of 0, 1, 2 or 3 to create NEWS (Box 1).[132]

Box 1 Chart for calculation of National Early Warning Score. Each vital sign is assigned a point sum based on the level of abnormality, the sum is the National Early Warning Score

Points	3	2	1	0	1	2	3
Respiratory Rate (/ minute)	≤8		9-11	12-20		21-24	≥25
Peripheral saturation (%)	≤91	92-93	94-95	≥96			
Use of supplemental oxygen		Yes		No			
Temperature (Celsius)	≤35.0		35.1-36.0	36.1-38.0	38.1-39	≥39.1	
Systolic blood pressure (mmHg)	≤90	91-100	101-110	111-219			≥220
Pulse (/minute)	≤40	41-50	41-50	51-90	91-110	111-130	≥131
Level of consciousness				A			V, P, or U

A, Alert; V, Verbal; P, Reacting to pain; U, Unconscious

Outcomes

The primary outcome was in-hospital mortality. The secondary outcome was Intensive Care Unit (ICU) admission. These outcomes were chosen for best possible comparison to the existing literature.[130] In the DMC, information regarding mortality and ICU admission was collected retrospectively from the Danish Civil Registration System and the Danish National Patient Register, respectively.[140, 141] In the NEED, outcome information was collected retrospectively from the Minimal Data Set.

Analysis

We combined NEWS with age in multivariable logistic regression to form NEWS+Age. Age was treated as a continuous predictor when fitted to the DMC development cohort. Prior to this, we assessed non-linearity of age in univariable logistic regression and explored non-linear terms for best fit. These included Restricted Cubic Splines (RCS) with 3 to 5 knots, age linear after age 40, age^2 after age 45, age+age^2, age^2 and logarithmic. Transformation terms were compared using the model Likelihood Ratio

χ^2 (See supplementary file 3). Ultimately, we transformed age based on RCS transformation with 4 knots. RCS knots were placed based on quantiles of the age distribution (knots at 24, 54, 71 and 88), as previously recommended.[142] The final number of knots was determined by maximizing the Akaike Information Criterion, penalizing more complex models (See supplementary file 3).[142]

We did not include non-linear terms for NEWS or interaction-terms between age and NEWS as these terms were insignificant when comparing the relative importance of terms in an ANOVA table of adjusted Wald χ^2 statistics. We applied the model coefficients and intercepts derived in DMC for assessment of predictive performance in the NEED.

For missing data in the DMC, we substituted missing vital sign measurements by multiple imputation to retain as much information as possible for model building. We used the chained equations procure,[143] after determining if imputation was feasible based on patterns of missingness.[142, 144] All vital signs in NEWS, age, sex, specific ED and both outcomes were included in the imputation procedure. We obtained 20 estimates of the missing vital sign for each patient. We averaged regression coefficients and intercepts to incorporate variance introduced by the imputation procedure. Predictive performance in the DMC was also based on the average estimate of risk.

A sensitivity analysis for impact of imputations on model fit was performed by comparing model estimates in the DMC data for both patients with missing variables imputed (i.e. *completed*) and patients with all variables available (i.e. *complete*).

For validation, we applied the model coefficients and intercepts derived in the DMC development cohort when assessing predictive performance in the NEED external validation cohort. Similarly, we assessed model performance in three age categories 18-64, 65-80 and above 80 in both the DMC and the NEED. Age-categories were pre-specified for comparison to previous studies.[42] We assessed discrimination, calibration, and net benefit. Discrimination was assessed by extracting AUROC. Calibration was assessed by extracting calibration slope and intercept from plots of observed outcomes by predicted risk and by visual inspection of flexible calibration curves.[145]

Net benefit was assessed using decision curve analysis, plotting the proportion of true cases identified while identifying no controls across a range of plausible high-risk thresholds. The difference between the curves represent the percentage increase compared to the other model.[146] The chosen risk-range correspond to the range between the median observed risk in the DMC with 3 NEWS points, to the median observed risk for persons with 7 NEWS points. In current clinical practice, 5 or 6 NEWS points trigger urgent clinical review.

The sample size was estimated to minimize model overfitting and to target sufficiently precise model predictions. The estimate was based on recommendations specific to prediction modelling using the pmsampsize package in R.[147] Based on previous observations in a similar population,[72] we set the expected outcome proportion conservatively to 0.024 for in-hospital mortality, set an anticipated $R^2_{\text{Cox-Schnell}}$ (i.e. variability explained by the model) of 0.18, with 25 predictor parameters in the model, assuming a 5% margin of error in estimation of model-intercept. This necessitated at least 2,237 subjects.

All analyses were performed in R statistical software.

RESULTS

Characteristics of Study Subjects

Table 1 provides a summary of patient characteristics for the DMC cohort. The median age of the 14,809 patients was 63 (IQR: 45-76) and 48.1% were women. A total of 365 patients (2.5%) died during the hospital stay and 417 (2.8%) were admitted to the ICU. In comparison, the median age of 50,448 patients in the NEED was 66 (IQR: 51-77), 1,413 (2.8%) patients died in-hospital and 824 (1.6%) were admitted to the ICU (See supplementary file 1). We excluded 2,039 patients from the DMC of whom 58 (2.8%) died in-hospital and 60 (2.9%) were admitted to the ICU (See supplementary file 2).

Table 1 Characteristics of patients in the Danish Multicenter Cohort used for development

Variable	**Age 18-64** (N=7,990)	**Age 65-80** (N=4,316)	**Age above 80** (N=2,503)	**All** (N=14,809)
Age, Median [IQR]	46 [34-57]	73 [69-77]	85 [83-89]	63 [45-76]
Female sex, No. (%)	3,804 (47.6%)	2,170 (50.3%)	1,151 (46.0%)	7,125 (48.1%)
Systolic blood pressure (mmHg)				
Median [IQR]	134 [121-147]	139 [123-156]	138 [120-157]	135 [121-151]
Missing, No. (%)	21 (0.3%)	11 (0.3%)	12 (0.5%)	44 (0.3%)
Pulse (/min)				
Median [IQR]	83 [71-97]	82 [70-96]	79 [68-93]	82 [70-96]
Missing, No. (%)	54 (0.7%)	40 (0.9%)	29 (1.2%)	123 (0.8%)
Respiratory Rate (/min)				
Median [IQR]	16 [14-19]	16 [15-20]	18 [16-24]	16 [14-20]
Missing, No. (%)	2,340 (29.3%)	1,073 (24.9%)	624 (24.9%)	4,037 (27.3%)
Peripheral saturation				
Median [IQR]	98 [96-99]	96 [94-98]	96 [94-98]	97 [95-98]
Missing, No. (%)	343 (4.3%)	159 (3.7%)	135 (5.4%)	637 (4.3%)
Level of consciousness, No. (%)				
Alert	6,891 (86.2%)	3,712 (86.0%)	2,039 (81.5%)	12,642 (85.4%)
Vocal	101 (1.3%)	84 (1.9%)	84 (3.4%)	269 (1.8%)

Table 1 Characteristics of patients in the Danish Multicenter Cohort used for development (continued)

Arial	**Age 18-64** (N=7,990)	**Age 65-80** (N=4,316)	**Age above 80** (N=2,503)	**All** (N=14,809)
Pain	36 (0.5%)	23 (0.5%)	22 (0.9%)	81 (0.5%)
Unresponsive	23 (0.3%)	10 (0.2%)	8 (0.3%)	41 (0.3%)
Missing, No (%)	939 (11.8%)	487 (11.3%)	350 (14.0%)	1,776 (12.0%)
Temperature (Celcius)				
Median [IQR]	37.0 [36.6-37.4]	36.9 [36.5-37.4]	36.9 [36.5-37.4]	37.0 [36.6-37.4]
Missing, No. (%)	526 (6.6%)	234 (5.4%)	182 (7.3%)	942 (6.4%)
Use of supplementary oxygen, No. (%)	353 (4.4%)	427 (9.9%)	309 (12.3%)	1,089 (7.4%)
NEWS				
Median [IQR]	1 [0-2]	2 [0-4]	2 [1-5]	1 [0-3]
Missing, No. (%)	3,128 (39.1%)	1,567 (36.3%)	982 (39.2%)	5,677 (38.3%)
In-hospital mortality, No. (%)	46 (0.6%)	156 (3.6%)	163 (6.5%)	365 (2.5%)
Admitted to Intensive Care Unit, No. (%)	197 (2.5%)	144 (3.3%)	76 (3.0%)	417 (2.8%)

IQR, Inter Quartile Range; DMC, Danish Multicenter Cohort; NEWS, National Early Warning Score

7

Main Results

In the DMC, age showed a non-linear association with the risk of in-hospital mortality or ICU admission in univariable analysis (See supplementary file 3 and 4). In multivariable analysis, age was the strongest predictor compared to NEWS sum-score for in-hospital mortality but was not an independent predictor for ICU admission (p: 0.58) (See supplementary file 5).

NEWS+Age showed good apparent validity in the DMC for prediction of in-hospital mortality with a significantly improved overall AUROC of 0.87 (95% CI: 0.85-0.88) compared to 0.82 (95% CI: 0.80-0.84) for NEWS (DeLong: $Z = -5.76$, $p < 0.001$) (Table 2 and Figure 1). Apparent overall calibration was nearly perfect as both NEWS and NEWS+Age were fitted to the DMC data. Within age categories, risk was systematically overestimated for the age category 18 to 64 with calibration slope increasing from 1.16 to 1.26, a worsening from the ideal value of 1 (Table 2). Flexible calibration curves suggested improved calibration for all age categories at risk-levels below 20% but also indicated a very low outcome incidence at risk levels above 20% (See supplementary file 6). In external validation of NEWS+Age, we found an AUROC of 0.82 (95% CI: 0.81-0.83) with overall satisfactory calibration. However, calibration slope increased from 1.10 to 1.17 for the age category 18-64 in the NEED, suggesting overestimation of risks (Table 2).

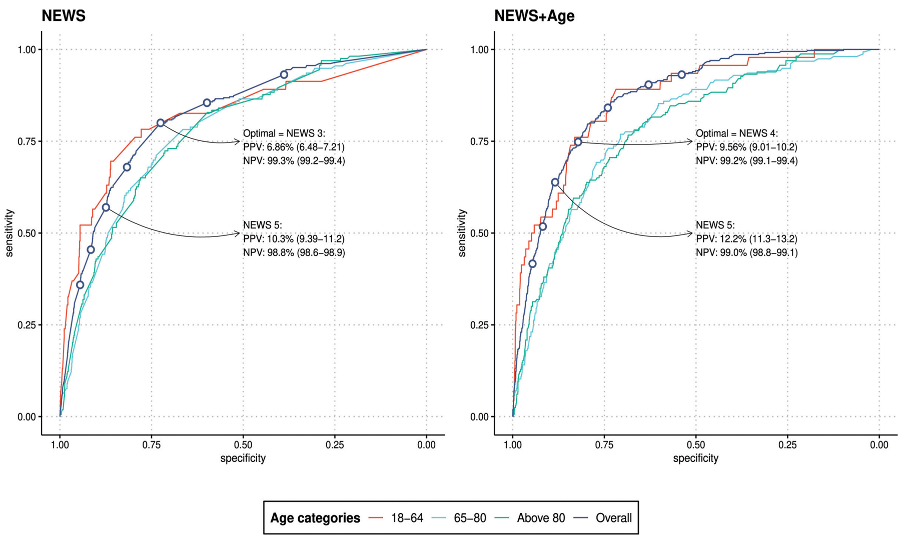


Figure 1 Receiver operating characteristics for NEWS and NEWS+Age by age groups and overall in the Danish Multicenter Cohort for prediction of in-hospital mortality. Values in parentheses indicate the 95% confidence intervals.

Abbreviations: NEWS, National Early Warning Score; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Table 2 Predictive performance for in-hospital mortality in the Danish Multicenter Cohort (Development) and the Netherlands Emergency Department evaluation cohort (External validation)

	Calibration		Discrimination			
	Intercept	**Slope**	**AUROC**		**95% CI**	
NEWS for in-hospital mortality in the DMC						
18-64	-1.26	1.16	0.82	0.75	-	0.89
65-80	0.24	0.87	0.78	0.74	-	0.81
Above 80	0.75	0.90	0.78	0.74	-	0.81
Overall	0.00	1.04	0.82	0.80	-	0.84
NEWS+Age for in-hospital mortality in the DMC						
18-64	0.02	1.26	0.87	0.81	-	0.92
65-80	0.05	0.99	0.79	0.75	-	0.82
Above 80	-0.05	1.03	0.78	0.75	-	0.82
Overall	0.00	1.03	0.87*	0.85	-	0.88
NEWS in the NEED						
18-64	-0.90	1.10	0.80	0.76	-	0.80
65-80	0.13	0.86	0.75	0.73	-	0.76
Above 80	0.69	0.74	0.72	0.70	-	0.74
Overall	-0.03	0.93	0.77	0.75	-	0.77
NEWS+Age for in-hospital mortality in the NEED						
18-64	0.36	1.17	0.83*	0.81	-	0.86
65-80	-0.04	0.94	0.75	0.73	-	0.77
Above 80	-0.10	0.85	0.73	0.71	-	0.75
Overall	0.00	0.90	0.82*	0.81	-	0.83

CI, Confidence Interval; AUROC, Area Under Curve for Receiver Operating Characteristic; NEWS, National Early Warning Score, DMC: Danish Multicenter Cohort; NEED, Netherlands Emergency Department Evaluation Database
* A significant difference was detected using DeLong's test compared to NEWS in the same cohort and age category.

NEWS is readily combined with age at the bedside by use of the conversion-chart (Figure 2). For example, a 76-year-old with a NEWS of 4 has a risk of in-hospital mortality below 6%, whereas an 84-year-old has a risk of 6-8% with the same NEWS. If 6% was considered a high-risk threshold for triggering, for example, urgent clinical review, the latter would trigger such a review, and all persons aged above 90 with a NEWS higher than 3 would trigger the same. Internal and external calibration of the conversion chart is presented

7

in Figure 3, confirming good calibration when used on the age category above 80 in both the DMC and the NEED.

Decision curves indicate that a decision-policy guided by NEWS+age is likely to increase the proportion of true cases to controls across the range of high-risk thresholds (See supplementary file 7).

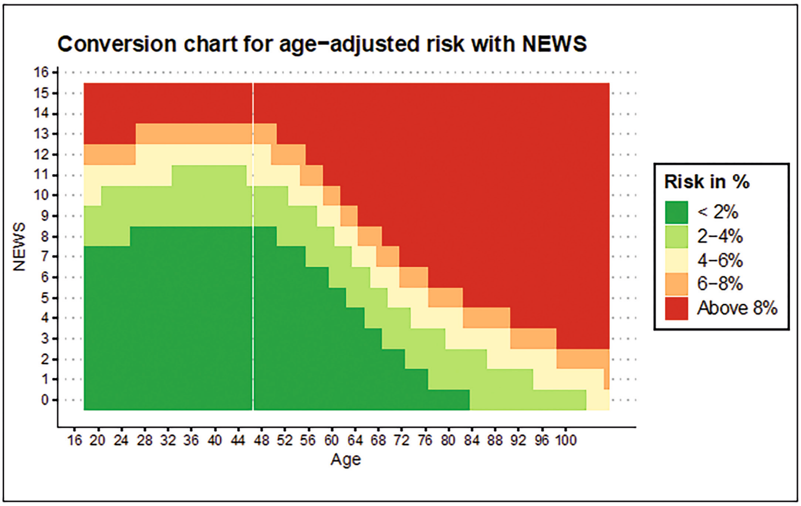


Figure 2 Conversion-chart for calculation of risk of in-hospital mortality with NEWS and age for for risk-categories <2%, 2-4%, 4-6%, 6-8% and >8%. Using this conversion chart, a 76-year-old carries a risk below 6% and an 84-year-old carries a risk of 6-8% with the same NEWS of 4.

Multiple imputation procedures reduced uncertainties in the estimates when fitted on multiply imputed data compared to complete data with only minor changes to model-coefficients (See supplementary file 5).

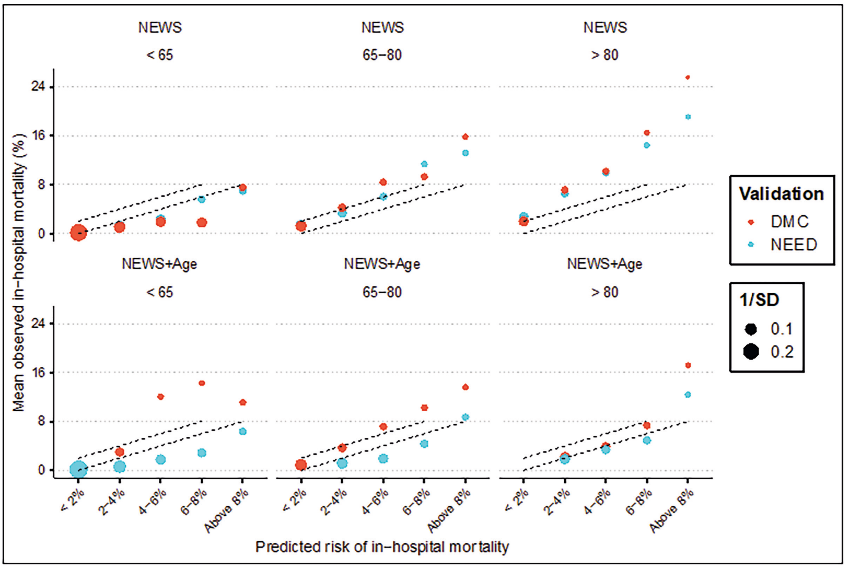


Figure 3 Internal and external calibration of the conversion-chart. The predicted risk of in-hospital mortality has been categorized to categories <2%, 2-4%, 4-6% and 6-8 in the conversion chart. Dots placed within dotted lines represent ideal calibration. The size of the dots indicate the precision of the estimate for observed in-hospital mortality in each risk-group, the larger, the higher the precision and is based on the inverse of the standard deviation.

Abbreviations: NEWS, National Early Warning Score; DMC, Danish Multicenter Cohort; NEED, Netherlands Emergency Department Evaluation Database

LIMITATIONS

An important limitation of our study is the risk of selection bias. We excluded patients from analyses in the DMC if both systolic blood pressure and pulse measurements were missing and these persons had slightly less favorable outcomes. We believe this bias will affect predictions mostly for persons at very high risk or very low risk with minor effects to our interpretation. Also, we performed external validation on patients with complete data in the NEED. As in the DMC, those with missing values likely differ, but by using complete data we perform an external validation that indicate the performance clinicians will experience unless they change habits.

We used multiple imputation procedures to replace missing values in the DMC dataset as per recent recommendations on reporting of EWS.[130] However, it is not possible to test our assumptions about patterns of

7

missingness, though only minor effects on our estimates would be expected because predictors included in both NEWS and NEWS+Age are well-associated to the outcome and the selection is comprehensive.[148] If we used complete data in our analysis, the amount of information for model fitting would have been considerably smaller and likely bias model estimates more.[149]

New models will often be favored when compared to unadjusted models in new settings.[150] NEWS is an integer score and as such provide the risk-estimate directly, making it impossible to adjust coefficients or intercept. However, in NEWS coefficients are equal for all vital signs after points have been awarded, and NEWS may as such be expected to fit linearly in new populations. We adjusted the estimated risks based on the best possible fit to the DMC data.

Treatment preference is likely an important factor we cannot account for in this study. The presence of advance directives may affect both in-hospital mortality and ICU admissions. It is fair to expect differences in prevalence of such directives when comparing Danish and Dutch populations and to expect advance directives to be most prevalent among the old, contributing to underestimation of risks for the oldest. However, we show similar findings of increased predictive performance of NEWS+Age for in-hospital mortality in both health care settings, suggesting limited impact of such directives for this outcome.

Finally, we observe considerable miscalibration for those persons with predicted risks above 20%. This groups consist of 215 persons in the DMC, corresponding to 1.4% of all, but we do not consider it plausible that a risk of in-hospital mortality above 20% would be noted without having triggered an action beforehand.

DISCUSSION

In this large international multicenter retrospective cohort study, we show that NEWS+Age improved prediction of in-hospital mortality but that the same model did not improve prediction of ICU admissions. Prediction of in-hospital mortality within age-categories revealed crucial differences: A systematic and severe underestimation of risks for persons aged above 80 was eliminated, while overestimation of risks for persons aged 18-64 remained. Predictions were shifted to slight overestimation for the age category 65-80. We validated our findings using external data in a different European country and different time-period, increasing generalizability considerably. We adhere to the TRIPOD guideline and use a very large sample sizes in relation to the number of prognostic variables to improve the validity of our estimates.

Our finding of increased discrimination for in-hospital mortality with NEWS+Age, is in line with studies from the ED setting using the similar Modified EWS,[133, 135] the prehospital setting,[42] and the general ward setting.[151] In comparison to sepsis patients in another study, we found much better discrimination with NEWS alone.[152] It is unsurprising as this subpopulation characterized by high suspicion of sequential organ-failure had higher NEWS on average and much higher in-hospital mortality compared to our all-comer population for which NEWS was developed. Additionally, a smaller sample increases the likelihood of extreme observations.

The treatment paradox has likely affected our findings. This bias rests on the assumption that those at high risk receive aggressive treatment, leading to apparent overestimation of risks because the outcome is often averted, a feature indicated previously.[153] Conversely, for the oldest age category, we found underestimation of risks which may be due to other unmeasured factors such as frailty, resuscitation preference or other limitations to treatment. We cannot exclude that ageism is present, as has also been suggested in the pre-hospital setting, where older patients are less likely to be transported to a higher level trauma-center.[154] Frailty has been suggested as an alternative to age in addition to vital signs at arrival to the ED,[155] and may add to measures of illness acuity for prediction of in-hospital mortality,[156] and post-admission mortality.[157, 158] Importantly, it should not be forgotten that frailty may itself become an aspect of ageism due to its perceived chronic nature,[159] and should demonstrate clear performance gains compared to our findings before implementation. For now, we demonstrate that underestimation of risks for older persons at arrival can be eliminated with NEWS+Age, and our findings may help ensure adequate levels of alertness until a clinical review can be performed.

Unexpectedly, we did not find NEWS+age to add meaningfully to prediction of ICU admissions compared to NEWS and therefore did not investigate predictive performance for ICU admission. An explanation may be that ICU admission is both treatment and outcome, in contrast to in-hospital mortality, and is less likely to happen for the oldest age categories in our cohorts (Table 1 and supplementary file 1). This is consistent with the treatment paradox and supports our finding that we cannot use one prediction model to predict both ICU admission and in-hospital mortality. Simulation studies suggests inclusion of treatment in prediction modelling could improve estimates of performance.[160] However, due to the differences in ICU admission proportions between age categories, this would worsen the observed treatment paradox. Also, ICU admission is not an available predictor at arrival where the prediction is made. Research from the intensive care setting now points to prognostic factors superior to chronological age that better capture physiological age, such as frailty, for prediction of ICU length of stay and mortality.[161] Mobility is another promising factor, readily available on arrival, to improve NEWS,[162, 163] but this was not recorded. Finally, treatment preferences may also be a prognostic factor to account for, but such preferences are often not present at arrival.

Discrimination as assessed by AUROC is often insensitive to addition of predictors,[164] and calibration is now considered crucial in evaluation of risk scores.[165] Our findings support this notion and it may explain why age was not included during development of NEWS.[132] Despite improving discrimination for persons aged 18-64 in external data, overestimation remained within the clinically relevant levels of risk. This overestimation could be reduced by varying the intercept of NEWS+Age by age category. For example, when building a prediction model for early death after traumatic bleeding, intercepts were adjusted to compensate for different baseline risk in high and low-income countries.[166] Another approach is determining age-dependent thresholds for each vital sign in NEWS,[44] but this requires a computer for practical application and the resulting model lacks the dissemination of NEWS. However, from Supplementary file 4 it is evident that age does not add much information for the age category below 65, indicating that only minor gains can be made. Instead, we sought to maintain NEWS in its original form and update it with chronological age, focusing on simplicity and non-linearity. This approach has also been recommended previously, due to a general tendency of reported models that are overfitted when applied in real-life.[20]

NEWS+Age may improve the ability to reliably determine risk of in-hospital mortality at arrival and act on time sensitive illnesses, especially if we apply it to persons aged above 80 in the ED. As the applicability to this sub-population is improved it may also increase appropriate use of NEWS.[167] We demonstrate the importance of considering baseline risks in different age-categories and a simple conversion chart to account for this. It is well established that without the statistical baseline-risk clearly presented, even statistical experts are biased in their perception of risk.[168] However, any modification to a score administered to all upon arrival is likely to have major implications on the workload for clinicians. On average, NEWS+Age is unlikely to increase the workload for clinicians due to improvements in both true positives and true negatives for the plausible high-risk-thresholds. However, randomized clinical trials testing the effect of NEWS+Age on both clinical decisions and outcome proportions are necessary. Prior to this, actual thresholds for considering persons high-risk should be decided after thorough discussion of clinically acceptable risks and resources available.[146]

In summary, combining NEWS with age increased discrimination significantly for in-hospital mortality and retained good overall calibration in external validation. Predictions improved considerably for persons aged above 80 but were overestimated slightly for persons aged below 80. Pending clinical trials, it could potentially improve clinical decision policies guided by the outcome in-hospital mortality.

SUPPLEMENTARY FILE 1

Characteristics of patients in the Netherlands Emergency Department evaluation cohort used for external validation

	Age 18-64 (N=25,071)	Age 65-80 (N=16,992)	Age above 80 (N=8,385)	All (N=50,448)
Age, Median [IQR]	51 [37-59]	73 [70-77]	85 [83-88]	66 [51-77]
Female sex, No. (%)	12,352 (49.3%)	7,472 (44.0%)	4,604 (54.9%)	24,428 (48.4%)
Systolic blood pressure (mmHg), Median [IQR]	128 [109-147]	137 [113-158]	141 [117-162]	132 [111-154]
Pulse (/min), Median [IQR]	85 [73-99]	84 [71-99]	82 [70-96]	84 [72-98]
Respiratory Rate (/min), Median [IQR]	16 [14-19]	18 [15-21]	18 [16-22]	17 [14-20]
Peripheral saturation (%), Median [IQR]	98 [97-100]	97 [95-99]	97 [95-99]	98 [96-99]
Level of consciousness, No. (%)				
Alert	24,853 (99.1%)	16,859 (99.2%)	8,266 (98.6%)	49,978 (99.1%)
Vocal	125 (0.5%)	78.0 (0.5%)	83.0 (1.0%)	286 (0.6%)
Pain	66.0 (0.3%)	34.0 (0.2%)	26.0 (0.3%)	126 (0.2%)
Unresponsive	27.0 (0.1%)	21.0 (0.1%)	10.0 (0.1%)	58.0 (0.1%)
Temperature (Celsius), Median [IQR]	36.9 [36.5-37.4]	36.8 [36.4-37.4]	36.8 [36.4-37.3]	36.9 [36.5-37.4]
Use of supplementary oxygen, No. (%)	1,085 (4.3%)	1,662 (9.8%)	1,160 (13.8%)	3,907 (7.7%)
NEWS, Median [IQR]	1 [0-3]	2 [0-4]	2 [1-5]	2 [0-4]
In-hospital mortality, No. (%)	260 (1.0%)	610 (3.6%)	543 (6.5%)	1413 (2.8%)
Admission to intensive care unit				
Yes, No. (%)	444 (1.8%)	305 (1.8%)	75.0 (0.9%)	824 (1.6%)
Missing, No. (%)	243 (1.0%)	88.0 (0.5%)	31.0 (0.4%)	362 (0.7%)

IQR, Inter Quartile Range; NEED, Netherlands Emergency Department Evaluation Database; NEWS, National Early Warning Score

SUPPLEMENTARY FILE 2

Characteristics of those excluded from the Danish Multicenter Cohort due to both systolic blood pressure and pulse measurements missing at arrival

Subjects excluded from the DMC cohort	Age 18-64 (N=989)	Age 65-80 (N=591)	Aged above 80 (N=459)	All (N=2039)
Age, Median [IQR]	46 [34-57]	73 [69-77]	86 [83-89]	67 [47-79]
Female sex, No. (%)	506 (51.2%)	337 (57.0%)	211 (46.0%)	1054 (51.7%)
In-hospital mortality, No. (%)	8 (0.8%)	23 (3.9%)	27 (5.9%)	58 (2.8%)
Admitted to Intensive Care Unit, No. (%)	24 (2.4%)	26 (4.4%)	10 (2.2%)	60 (2.9%)
NEWS variables missing *, Mean [Min-Max]	6.86 [4-7]	6.90 [4-7]	7.93 [4-7]	6.89 [4-7]

IQR, Inter Quartile Range; NEWS, National Early Warning Score; DMC, Danish Multicenter Cohort
* There are seven predictors in NEWS

SUPPLEMENTARY FILE 3

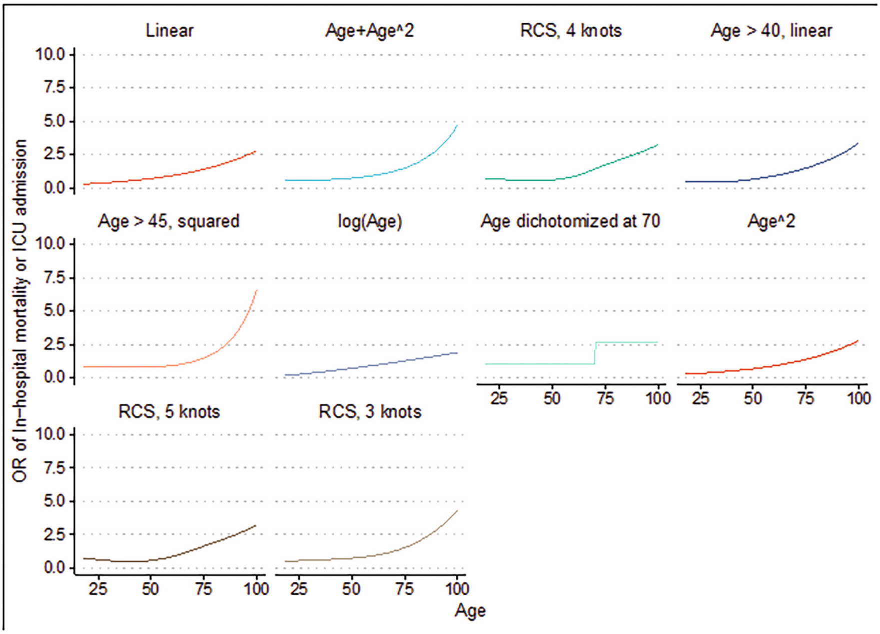
Univariate regression fits of age to ICU admission or in-hospital mortality according to transformation-term and derived measures of fit

Transformation term	Degrees of freedom	LR- statistic	Nagelkerkes' R2	AIC	Information loss by LR-statistic (%)
RCS, 4 knots	3	193	41	5502	0
RCS, 5 knots	4	193	41	5504	0
Age>40, linear	1	188	39	5503	3
Age+Age^2	2	183	39	5510	5
RCS, 3 knots	2	183	39	5510	5
Linear	1	172	36	5519	11
Age^2	1	172	36	5519	11
(Age>45)^2	1	169	36	5522	13
Age, dichotomized at 70	1	157	33	5534	19
Log(Age)	1	145	31	5546	25

LR, Likelihood Ratio; AIC, Akaikes Information Criterion; RCS, Restricted Cubic Splines

SUPPLEMENTARY FILE 4

A visual plot of odds-ratios to age depending on type of age-transformation tested. Odds ratios are compared to the odds at the median age of 64 in the Danish Multicenter Cohort Abbreviations: *RCS,* restricted cubic splines regression; *OR,* Odds Ratio



SUPPLEMENTARY FILE 5

Sensitivity analysis for impact of imputation procedures on model fit for in-hospital mortality in the DMC cohort

	Coefficient	S.E.	Pr(>|Wald Z|)	AUROC
NEWS+Age for **in-hospital mortality**: Fitted on completed cases (missing values imputed)[b]				
RCS(Age)[a]	1.86	0.145	<0.001	0.87 (95% CI: 0.85-0.88)
NEWS	0.29	0.016	<0.001	
NEWS+Age for **in-hospital mortality**: Fitted on complete cases (missing values omitted)				
RCS(Age)[a]	1.88	0.191	<0.001	0.86 (95% CI: 0.84-0.88)
NEWS	0.28	0.020	<0.001	
NEWS+Age for **ICU admission**: Fitted on completed cases (missing values estimated)				
RCS(Age)[a]	-0.05	0.095	0.584	0.67 (95% CI: 0.64-0.70)
NEWS	0.23	0.015	<0.001	

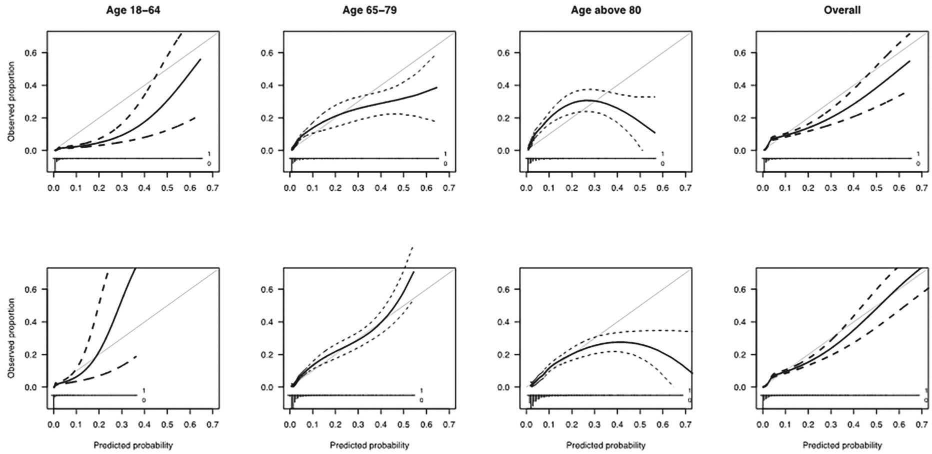
Abbreviations: S.E., Standard Error; NEWS, National Early Warning Score; RCS, Restricted Cubic Splines; Pr, Probability; AUROC, Area Under Receiver Operating Characteristic; CI, Confidence Interval.

[a] Formula: -3.33-0.014 * Age+0.000019 * pmax(Age-24)3 -0.000077 * pmax(Age-54)3 + 0.000085 * pmax(Age-71)3 -0.000026 * pmax(age-88)3

[b] Formula: 0.59 + 0.29 * NEWS + 1.86 * RCS(age)

SUPPLEMENTARY FILE 6

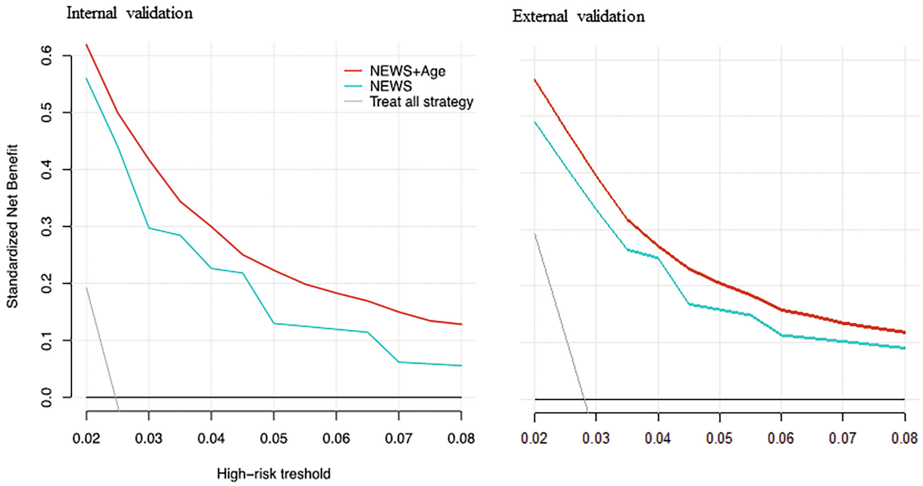
Flexible calibration plots for NEWS and NEWS+Age by age groups and overall, in the DMC. Bold lines are flexible calibration fits. A non-parametric loess smoother is used instead of restricted cubic in age categories 65-80 and above 80 due to number of observations being below 5000. Grey lines represent ideal calibration. Dotted lines denote 95% confidence interval of the flexible calibration curve. Considerable miscalibration is present for patients with predicted risks above 20%. This groups consist of 215 patients in the DMC, corresponding to 1.4% overall. It is not plausible that a risk of in-hospital mortality above 20% would be noted without having triggered an action beforehand.



Abbreviations: NEWS, National Early Warning Score; DMC, Danish Multicenter Cohort.

SUPPLEMENTARY FILE 7

Decision curves comparing a decision policy guided by NEWS to a policy guided by NEWS+Age for a range of plausible high-risk thresholds in the DMC (A) and NEED (B). Standardized net benefit is the proportion of true cases identified while identifying 0 controls at a given high-risk threshold.



Abbreviations: NEWS, National Early Warning Score; DMC, Danish Multicenter Cohort; NEED, Netherlands Emergency Department Evaluation Database