The Impact of Age on Predictive Performance of National Early Warning Score at Arrival to Emergency Departments

Development and External Validation

Nissen, Søren K.; Candel, Bart G.J.; Nickel, Christian H.; de Jonge, Evert; Ryg, Jesper; Bogh, Søren B.; de Groot, Bas; Brabrand, Mikkel

Published in:
Annals of Emergency Medicine

DOI:
10.1016/j.annemergmed.2021.09.434

Publication date:
2022

Document version:
Final published version

Citation for pulished version (APA):
Nissen, S. K., Candel, B. G. J., Nickel, C. H., de Jonge, E., Ryg, J., Bogh, S. B., de Groot, B., & Brabrand, M. (2022). The Impact of Age on Predictive Performance of National Early Warning Score at Arrival to Emergency Departments: Development and External Validation. *Annals of Emergency Medicine*, *79*(4), 354-363. https://doi.org/10.1016/j.annemergmed.2021.09.434

The Impact of Age on Predictive Performance of National Early Warning Score at Arrival to Emergency Departments: Development and External Validation

Søren K. Nissen, MD*; Bart G. J. Candel, MD; Christian H. Nickel, MD, PhD; Evert de Jonge, MD, PhD; Jesper Ryg, MD, PhD; Søren B. Bogh, MSc, PhD; Bas de Groot, MD, PhD; Mikkel Brabrand, MD, PhD

*Corresponding Author. E-mail: sknissen@health.sdu.dk.

Study objective: To investigate how age affects the predictive performance of the National Early Warning Score (NEWS) at arrival to the emergency department (ED) regarding inhospital mortality and intensive care admission.

Methods: International multicenter retrospective cohorts from 2 Danish and 3 Dutch ED. Development cohort: 14,809 Danish patients aged ≥18 years with at least systolic blood pressure or pulse measured from the Danish Multicenter Cohort. External validation cohort: 50,448 Dutch patients aged ≥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 to 64 years, 65 to 80 years, and more than 80 years.

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

Conclusion: Combining NEWS with age improved the prediction performance regarding inhospital mortality, mostly for persons aged above 80, and can potentially improve decision policies at arrival to EDs. [Ann Emerg Med. 2022;79:354-363.]

Please see page 355 for the Editor's Capsule Summary of this article.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2021.09.434

SEE EDITORIAL, P. 364

INTRODUCTION

Background

Early warning scores aggregate vital signs to assess illness acuity and identify persons at risk of deterioration in the emergency department (ED).[1] Among early warning scores, the National Early Warning Score (NEWS) in particular has seen widespread implementation.[2] NEWS was developed after a thorough review of 33 other early warning scores and combines 6 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.[3] However, the ability to predict the prognosis of older ED patients using NEWS may be inadequate.[4,5] One explanation may be that the risk of mortality associated with increasing vital sign abnormalities is larger in older persons, or that vital signs discriminate worse for older compared to younger persons for a combined outcome of inhospital mortality or ICU admission.[6,7] In the development of NEWS, age was not included as a predictor because the impact of age on predictive performance was not considered convincing by developers.[3]

Importance

Older persons 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

Editor's Capsule Summary

What is already known on this topic

The National Early Warning Score (NEWS) is a prediction tool that uses 6 physiologic parameters to identify patients at high risk for clinical deterioration.

What question this study addressed

Does the addition of age improve the predictive accuracy of NEWS?

What this study adds to our knowledge

In 2 data sets (n=14,809 and 50,448), the accuracy of NEWS as a predictor of inhospital mortality increased with the addition of age.

How this is relevant to clinical practice

Emergency departments that use NEWS to provide clinicians with information about expected mortality ought to consider NEWS+Age. More information regarding the impact on clinical care is needed.

either dichotomized or categorized age when investigating how to combine age with the similar Modified Early Warning Scores.[6,8] However, chronological age may not always reflect physiological age, and risk may increase in a nonlinear fashion. Thus, the gain or harm from combining NEWS with age remains uncertain.

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 validate the predictive performance externally.

MATERIALS AND METHODS

Study Design and Setting

This was an international multicenter cohort study based on existing cohorts from EDs. Analysis of predictive performance was performed for preplanned subgroups, stratified by age to 3 categories (18 to 64 years, 65 to 80 years, and more than 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.[9]

Selection of Participants

Development cohort. We used data from the Danish Multicenter Cohort and combined NEWS with age. The Danish Multicenter Cohort includes patients aged ≥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 2 Level II emergency centers: University Hospital of Southwest Jutland: (October 2, 2008, to February 12, 2009; February 23, 2010, to May 26, 2010; June 1, 2012, to November 1, 2011; and April 24, 2013, to December 9, 2013), and Lillebaelt Hospital (January 1, 2010, to June 30, 2010).[10-13]

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 would include 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 were available from a Level III emergency center (January 1, 2017, to June 8, 2019) and 2 emergency centers at Level II (January 1, 2019, to January 12, 2020; and January 1, 2017, to December 31, 2019).

Methods of Measurement

In the Danish Multicenter Cohort, 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.[14] Each vital sign was then assigned a value of 0, 1, 2, or 3 to create NEWS (Table 1).[3]

Outcome Measures

The primary outcome was inhospital mortality. The secondary outcome was ICU admission. These outcomes were chosen for the best possible comparison to the existing literature.[1] In the Danish Multicenter Cohort, information regarding mortality and ICU admission was collected retrospectively from the Danish Civil Registration System and the Danish National Patient Register, respectively.[15,16] In the NEED, outcome information was collected retrospectively from the Minimal Data Set.

Primary Data 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 Danish Multicenter Cohort development cohort. Prior to this, we assessed the nonlinearity of age in univariable logistic regression and explored nonlinear terms for best fit. These included restricted cubic splines 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 (Table E1 [available at http://www.annemergmed.com]). Ultimately, we transformed age based on restricted cubic splines transformation with 4 knots. Restricted cubic splines knots were placed based on quantiles of the age distribution (knots at 24, 54, 71, and 88), as previously recommended.[17] The final number of knots was determined by maximizing the Akaike Information Criterion, penalizing more complex models (Table E1).[17]

We did not include nonlinear 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 the Danish Multicenter Cohort for the assessment of predictive performance in the NEED.

For missing data in the Danish Multicenter Cohort, we substituted missing vital sign measurements with multiple imputations to retain as much information as possible for model building. We used the chained equations procure, after determining if imputation was feasible based on patterns of missingness.[17-19] 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 Danish Multicenter Cohort was also based on the average estimate of risk.

A sensitivity analysis for the effect of imputations on the model fit was performed by comparing model estimates in the Danish Multicenter Cohort data for both patients with missing variables imputed (ie, "completed") and patients with all variables available (ie, "complete").

For validation, we applied the model coefficients and intercepts derived in the Danish Multicenter Cohort when assessing predictive performance in the NEED external validation cohort. Similarly, we assessed model performance in 3 age categories 18 to 64 years, 65 to 80 years, and more than 80 years in both the Danish Multicenter Cohort and the NEED. Age categories were prespecified for comparison to previous studies.[5] 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 visual inspection of flexible calibration curves.[20]

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 represents the percentage increase compared to the other model.[21] The chosen risk range corresponds to the range between the median observed risk in the Danish Multicenter Cohort with 3 NEWS points to the median observed risk for

Table 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 (breaths/min)	≤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 (mm Hg)	≤90	91-100	101-110	111-219			≥220
Pulse (beats/min)	≤40	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.

persons with 7 NEWS points. In current clinical practice, 5 or 6 NEWS points trigger an 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 modeling using the pmsampsize package in R.[22] Based on previous observations in a similar population, we set the expected outcome proportion conservatively to 0.024 for inhospital mortality, set an anticipated Cox & Schnell R^2 (ie, variability explained by the model) of 0.18, with 25 predictor parameters in the model, assuming a 5% margin of error in the estimation of model-intercept. This necessitated at least 2,237 subjects.[23]

All analyses were performed in R statistical software.[24]

RESULTS

Characteristics of Study Subjects

Table 2 provides a summary of patient characteristics for the Danish Multicenter Cohort. The median age of the 14,809 patients was 63 (interquartile range, 45 to 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 (interquartile range, 51 to 77), 1,413 (2.8%) patients died inhospital, and 824 (1.6%) were

Table 2. 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 (beats/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 (breaths/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)
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 (degrees 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)
Inhospital 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)

DMC, Danish Multicenter Cohort; *IQR*, interquartile range.

admitted to the ICU (Table E2 [available at http://www.annemergmed.com]). We excluded 2,039 patients from the Danish Multicenter Cohort, of whom 58 (2.8%) died inhospital, and 60 (2.9%) were admitted to the ICU (Table E3 [available at http://www.annemergmed.com]).

Main Results

In the Danish Multicenter Cohort, age showed a nonlinear association with the risk of inhospital mortality or ICU admission in univariable analysis (Figure E1 [available at http://www.annemergmed.com] and Table E1). In multivariable analysis, age was the strongest predictor compared to NEWS sum-score for inhospital mortality but was not an independent predictor for ICU admission (P=.58) (Table E4 [available at http://www.annemergmed.com]).

NEWS+Age showed good apparent validity in the Danish Multicenter Cohort for prediction of inhospital mortality with a significantly improved overall AUROC of 0.86 (95% confidence interval [CI] 0.85 to 0.88) vs 0.82 (95% CI 0.80 to 0.84) for NEWS (DeLong: $Z = -5.76$, $P<.001$) (Table 3 and Figure 1). Apparent overall calibration was nearly perfect as both NEWS and NEWS+Age were fitted to the Danish Multicenter Cohort data. Within age categories, the risk was systematically overestimated for the age category 18 to 64 years with calibration slope increasing from 1.16 to 1.26, a worsening from the ideal value of 1 (Table 3). 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% (Figure E2 [available at http://www.annemergmed.com]). In external validation of NEWS+Age, we found an AUROC of 0.82 (95% CI 0.80 to 0.84) with overall satisfactory calibration. However, the calibration slope increased from 1.10 to 1.17 for the age category 18 to 64 years in the NEED, suggesting overestimation of risks (Table 3).

NEWS is readily combined with age at the bedside by using a conversion chart (Figure 2). For example, a 76-year-old with a NEWS of 4 has a risk of inhospital mortality below 6%, whereas an 84-year-old has a risk of 6% to 8% with the same NEWS. If 6% were considered a high-risk threshold for triggering, for example, urgent clinical review, the latter would trigger such a review, and

Table 3. Predictive performance for inhospital mortality in the Danish Multicenter Cohort (Development) and the Netherlands Emergency Department evaluation cohort (External validation).

	Calibration		Discrimination			
Model and cohort	**Intercept**	**Slope**	**AUROC**	**95% CI**		
NEWS for inhospital 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 inhospital 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	0.86*	0.85	-	0.88
NEWS for inhospital mortality 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	0.75	0.75	-	0.77
NEWS+Age for inhospital 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

*A significant difference was detected using DeLong's test compared to NEWS in the same cohort and age category.

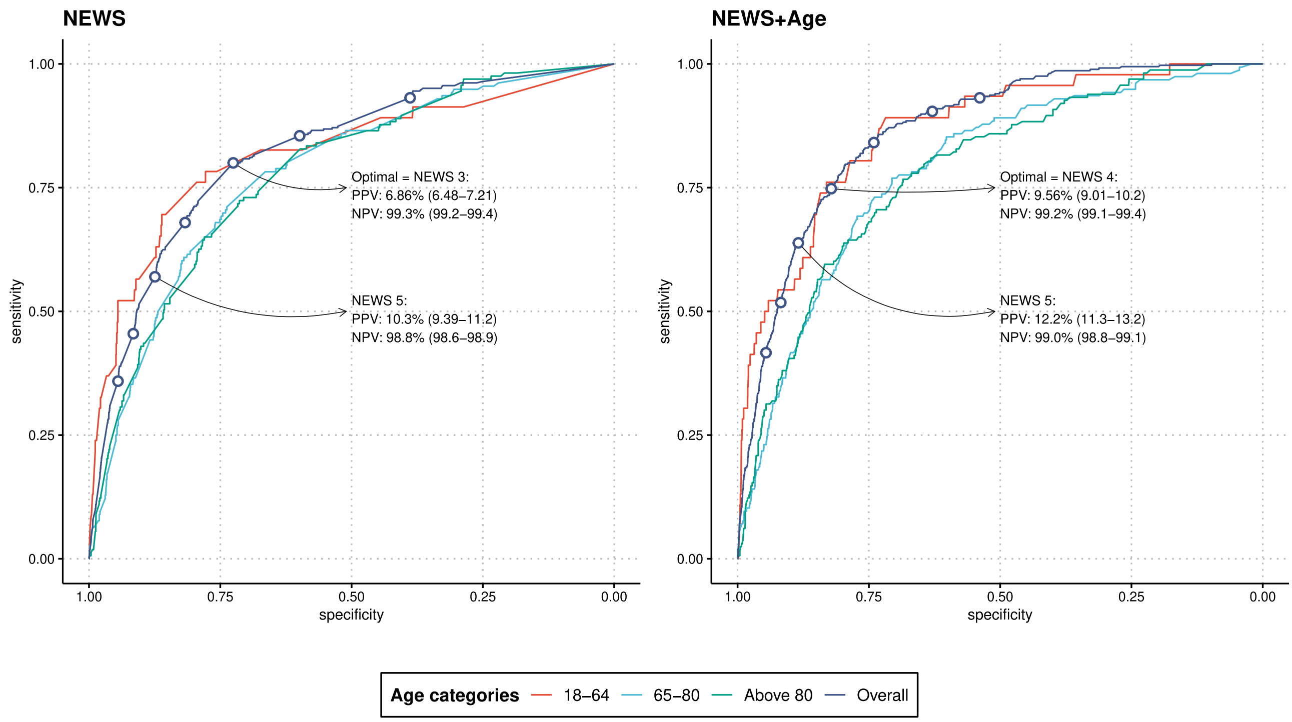

Figure 1. Receiver operating characteristics for NEWS and NEWS+Age by age groups and overall in the Danish Multicenter Cohort for prediction of inhospital mortality. Values in parentheses indicate the 95% confidence intervals. *NEWS,* National Early Warning Score; *PPV,* Positive Predictive Value; *NPV*, Negative Predictive Value.

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 in Figure 3, confirming good calibration when used on the age category above 80 in both the Danish Multicenter Cohort 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 (Figure E3 [available at http://www.annemergmed.com]).

Multiple imputation procedures reduced uncertainties in the estimates when fitted on multiply imputed data

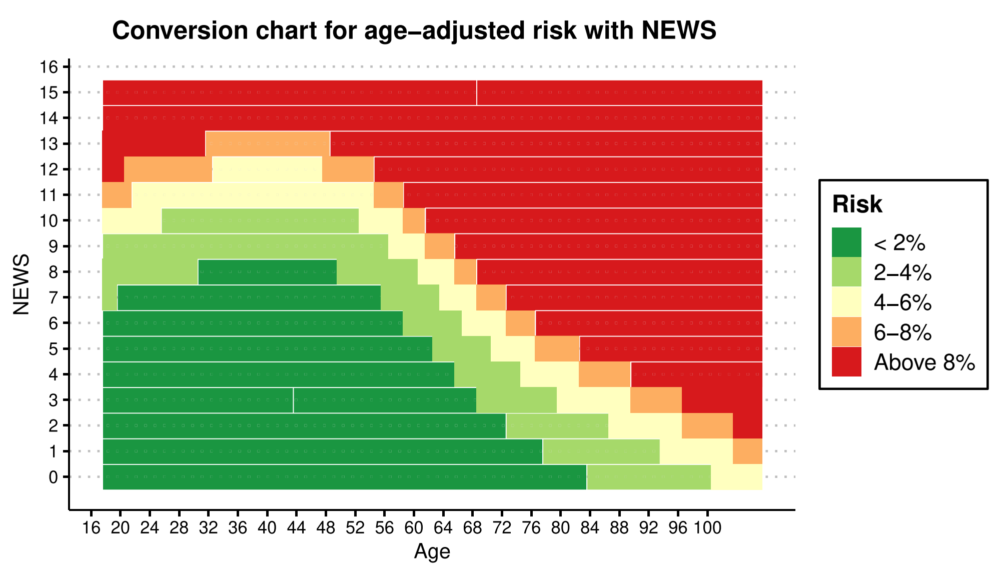

Figure 2. Conversion chart for calculation of risk of inhospital mortality with NEWS and age for risk categories less than 2%, 2% to 4%, 4% to 6%, 6% to 8%, and more than 8%. Using this conversion chart, a 76-year-old carries a risk below 6%, and an 84-year-old has a 6% to 8% risk with the same NEWS of 4.

compared with complete data with only minor changes to model coefficients (Table E4).

LIMITATIONS

An important limitation of our study is the risk of selection bias. We excluded patients from analyses in the Danish Multicenter Cohort 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 on our interpretation. Also, we performed external validation on patients with complete data in the NEED. As in the Danish Multicenter Cohort, those with missing values likely differ, but by using complete data, we perform an external validation that indicates the performance clinicians would experience unless they change habits.

We used multiple imputation procedures to replace missing values in the Danish Multicenter Cohort data set as per recent recommendations on reporting Early Warning Signs.[1] However, it is impossible to test our assumptions about patterns of 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.[25] 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.[26]

New models will often be favored when compared to unadjusted models in new settings.[27] 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 Danish Multicenter Cohort data.

Treatment preference is likely an important factor we cannot account for in this study. The presence of advance directives may affect both inhospital mortality and ICU admissions. It is fair to expect differences in the 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 the increased predictive performance of NEWS+Age for inhospital mortality in both health care settings, suggesting a limited impact of such directives for this outcome.

Finally, we observe considerable miscalibration for those persons with predicted risks above 20%. This group

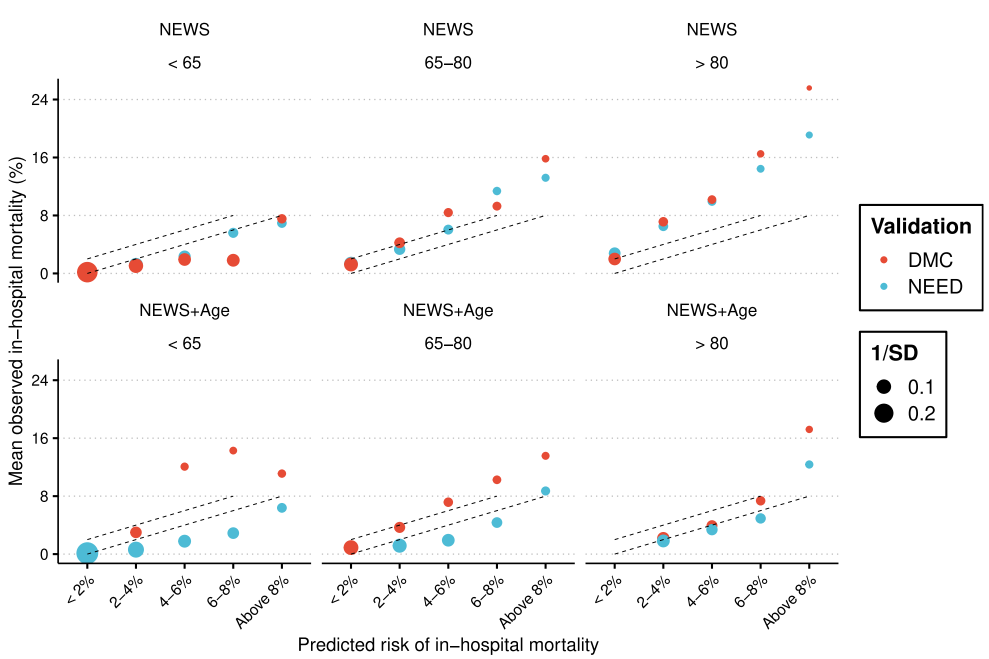

Figure 3. Internal and external calibration of the conversion chart. The predicted risk of inhospital mortality has been categorized into less than 2%, 2% to 4%, 4% to 6%, and 6% to 8% in the conversion chart. Dots placed within dotted lines represent ideal calibration. The size of the dots indicates the precision of the estimate for observed inhospital mortality in each risk group– the larger, the higher the precision–and is based on the inverse of the standard deviation. *DMC,* Danish Multicenter Cohort.

consists of 215 persons in the Danish Multicenter Cohort, corresponding to 1.4% of all, but we do not consider it plausible that the risk of inhospital 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 inhospital mortality but that the same model did not improve prediction of ICU admissions. Prediction of inhospital mortality within age categories revealed crucial differences: a systematic and severe underestimation of risks for persons aged more than 80 years was eliminated, while overestimation of risks for persons aged 18 to 64 years remained. Predictions were shifted to slight overestimation for the age category 65 to 80 years. 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 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 inhospital mortality with NEWS+Age aligns with studies from the ED setting using the similar Modified Early Warning Score, the out-of-hospital setting, and the general ward setting.[5,6,8,28] In comparison to sepsis patients in another study, we found much better discrimination with NEWS alone.[4] It is not surprising as this subpopulation characterized by high suspicion of sequential organ failure had higher NEWS on average and much higher inhospital 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.[29] 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 an out-of-hospital setting, where older patients are less likely to be transported to a higher-level trauma center.[30] Frailty has been suggested as an alternative to age in addition to vital signs at arrival to the ED, and may add to measures of illness acuity for prediction of inhospital mortality and postadmission mortality.[31-34] Importantly, it should not be forgotten that frailty may itself become an aspect of ageism due to its perceived chronic nature, and should demonstrate clear performance gains compared to our findings before implementation.[35] 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 a prediction of ICU admissions compared to NEWS. Therefore we did not investigate the predictive performance for ICU admission. An explanation may be that ICU admission is both treatment and outcome, in contrast to inhospital mortality, and is less likely to happen for the oldest age categories in our cohorts (Table 2 and Table E2). This is consistent with the treatment paradox and supports our finding that we cannot use one model to predict both ICU admission and inhospital mortality. Simulation studies suggest the inclusion of treatment in prediction modeling could improve estimates of performance.[36] 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.[37] Mobility is another promising factor, readily available on arrival, to improve NEWS[38,39]; however, 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.

As assessed by AUROC, discrimination is often insensitive to the addition of predictors, and calibration is now considered crucial in evaluating risk scores.[40,41] Our findings support this notion, and it may explain why age was not included during the development of NEWS.[3] Despite improving discrimination for persons aged 18 to 64 years 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 risks in high and low-income countries.[42] Another approach is determining age-dependent thresholds for each vital sign in NEWS[43]; however, this requires a computer for practical application, and the resulting model lacks the dissemination of NEWS. However, from Figure E1, 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 nonlinearity. 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 the risk of inhospital 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 subpopulation is improved, it may also increase the appropriate use of NEWS.[44] 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.[45] 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 as high-risk should be decided after a thorough discussion of clinically acceptable risks and resources available.[21]

In summary, combining NEWS with age increased discrimination significantly for inhospital 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 of inhospital mortality.

The authors acknowledge the participation of the nursing staff from the Emergency Departments in Denmark and recognize Line Emilie Lilholm Laugesen, MScPH, Mette Rahbek Kristensen, RN, Tanja Mose Kristensen, RN, and Anne-Mette Green, RN, for playing an integral role in coordinating data collection in Denmark.

Supervising editor: Timothy F. Platts-Mills, MD, MSc. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From the Institute of Regional Health Research (Nissen, Brabrand), Center South-West Jutland, University of Southern Denmark, Esbjerg, Denmark; the Department of Emergency Medicine (Nissen, Brabrand), Hospital of South-West Jutland, Esbjerg, Denmark; the Department of Emergency Medicine (Candel, Groot), Leiden University Medical Center, Leiden, the Netherlands; the Department of Emergency Medicine (Candel), Máxima Medical Center, Veldhoven, the Netherlands; the Department of Emergency Medicine (Nickel), University Hospital Basel, University of Basel, Basel, Switzerland; the Department of Intensive Care Medicine (de Jonge), Leiden University Medical Center, Leiden, the Netherlands; the Department of Clinical Research (Ryg), University of Southern Denmark, Odense, Denmark; the Department of Geriatric Medicine (Ryg), Odense University Hospital, Odense, Denmark; the Odense Patient Exploratory Network (OPEN) (Bogh), University of Southern Denmark, Odense, Denmark; and the Department of Emergency Medicine (Brabrand), Odense University Hospital, Odense, Denmark.

Author contributions: CHN, MB, and SKN conceived the study. SKN and BJGC drafted the protocol. All authors contributed to the study design. BdG and BJGC contributed to data collection in the Netherlands. MB and SKN contributed to data collection in Denmark. SKN, BJGC, and SBB performed the statistical analysis. All authors contributed to the interpretation of data analysis. SKN drafted the manuscript. All authors contributed substantially to revisions of the manuscript. All authors contributed to the critical revision of the manuscript and approved its final version. SKN takes responsibility for the manuscript as a whole.

Authorship: All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist. This study was supported by funding from the University of Southern Denmark and the Region of Southern Denmark. Funding bodies had no influence on study design, data collection, interpretation, or decision on publication.

Publication dates: Received for publication July 29, 2021. Revisions received September 1, 2021, and September 17, 2021. Accepted for publication September 24, 2021.

REFERENCES

1. Gerry S, Bonnici T, Birks J, et al. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. *BMJ*. 2020;369:m1501.
2. Royal College of Physicians. National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party; 2017. Accessed November 28, 2019. https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2
3. Royal College of Physicians. National Early Warning Score (NEWS): Standardising the assessment of acute-illness severity in the NHS. Report of a working party. 2012. Accessed November 28, 2019. https://www.rcplondon.ac.uk/file/9559/download

4. de Groot B, Stolwijk F, Warmerdam M, et al. The most commonly used disease severity scores are inappropriate for risk stratification of older emergency department sepsis patients: an observational multi-centre study. *Scand J Trauma Resusc Emerg Med*. 2017;25:91.
5. Pirneskoski J, Lääperi M, Kuisma M,et al. Ability of prehospital NEWS to predict 1-day and 7-day mortality is reduced in the older adult patients. *Emerg Med J*. Published online May 11, 2021. https://doi.org/10.1136/emermed-2019-209400
6. Smith GB, Prytherch DR, Schmidt PE, et al. Should age be included as a component of track and trigger systems used to identify sick adult patients? *Resuscitation*. 2008;78:109-115.
7. Lamantia MA, Stewart PW, Platts-Mills TF, et al. Predictive value of initial triage vital signs for critically ill older adults. *West J Emerg Med*. 2013;14:453-460.
8. Subbe CP, Kruger M, Rutherford P, et al. Validation of a modified early warning score in medical admissions. *QJM - Mon J Assoc Physicians*. 2001;94:521-526.
9. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162:55-63.
10. Brabrand M, Knudsen T, Hallas J, et al. The Paris score can reliably predict 7-day all-cause mortality for both acute medical and surgical patients: an international validation study. *Qjm*. 2018;111:721-725.
11. Nielsen MK, Henriksen DP, Østergaard KB, et al. Association between presenting complaints of acutely admitted medical patients and mortality: a cohort study. *Eur J Intern Med*. 2018;54:e29-e32.
12. 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.
13. Öhman MC, Atkins TEH, Cooksley T, et al. Validation of the MARS: a combined physiological and laboratory risk prediction tool for 5- to 7-day in-hospital mortality. *Qjm*. 2018;111:385-388.
14. De Hond A, Raven W, Schinkelshoek L, et al. Machine learning for developing a prediction model of hospital admission of emergency department patients: hype or hope? *Int J Med Inform*. 2021;152:104496.
15. Schmidt M, Schmidt SAJ, Sandegaard JL, et al. The Danish National patient registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449-490.
16. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol*. 2014;29:541-549.
17. Harrell FE. *Regression Modeling Strategies*. vol 45. Cham: Springer International Publishing; 2015.
18. Buuren S van, Groothuis-Oudshoorn K. mice : multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45(3):1-67.
19. Rubin DB. Inference and Missing Data. *Biometrika*. 1976;63:581-592.
20. Van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016;74:167-176.
21. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352:3-7.
22. Ensor J, Martin EC, Riley RD. pmsampsize: Calculates the Minimum Sample Size Required for Developing a Multivariable Prediction Model. 2019. Accessed June 10, 2020. https://cran.r-project.org/package=pmsampsize
23. Redfern OC, Pimentel MAF, Prytherch D, et al. Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: development and validation of a multivariable model. *Resuscitation*. 2018;133:75-81.
24. R Core Team. *A Language and Environment for Statistical Computing*. 2019. Accessed June 20, 2019. https://www.r-project.org
25. Heymans MW, Van Buuren S, Knol DL, et al. Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Med Res Methodol*. 2007;7:33.
26. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:157-160.
27. Collins GS, Moons KGM. Comparing risk prediction models: should be routine when deriving a new model for the same purpose. *BMJ*. 2012;344:1-2.
28. Lee YS, Choi JW, Park YH, et al. Evaluation of the efficacy of the National Early Warning Score in predicting in-hospital mortality via the risk stratification. *J Crit Care*. 2018;47:222-226.
29. De Groot B, Lameijer J, De Deckere ERJT, et al. The prognostic performance of the predisposition, infection, response and organ failure (PIRO) classification in high-risk and low-risk emergency department sepsis populations: comparison with clinical judgement and sepsis category. *Emerg Med J*. 2014;31:292-300.
30. Garwe T, Stewart K, Stoner J, et al. Out-of-hospital and inter-hospital under-triage to designated tertiary trauma centers among injured older adults: a 10-year statewide geospatial-adjusted analysis. *Prehospital Emerg Care*. 2017;21:734-743.
31. Schultz M, Rasmussen LJH, Carlson N, et al. Risk assessment models for potential use in the emergency department have lower predictive ability in older patients compared to the middle-aged for short-term mortality-a retrospective cohort study. *BMC Geriatr*. 2019;19:1-9.
32. Romero-Ortuno R, Wallis S, Biram R, Keevil V. Clinical frailty adds to acute illness severity in predicting mortality in hospitalized older adults: an observational study. *Eur J Intern Med*. 2016;35:24-34.
33. Kaeppeli T, Rueegg M, Dreher-Hummel T, et al. Validation of the clinical frailty scale for prediction of thirty-day mortality in the emergency department. *Ann Emerg Med*. 2020;76:291-300.
34. Engvig A, Wyller TB, Skovlund E, et al. Association between clinical frailty, illness severity and post-discharge survival: a prospective cohort study of older medical inpatients in Norway. *Eur Geriatr Med*. Published online August 21, 2021. https://doi.org/10.1007/s41999-021-00555-8
35. Dent E, Martin FC, Bergman H, Woo J, Romero-Ortuno R, Walston JD. Management of frailty: opportunities, challenges, and future directions. *Lancet*. 2019;394:1376-1386.
36. Groenwold RHH, Moons KGM, Pajouheshnia R, et al. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. *J Clin Epidemiol*. 2016;78:90-100.
37. Guidet B, de Lange DW, Boumendil A, et al. The contribution of frailty, cognition, activity of daily life and comorbidities on outcome in acutely admitted patients over 80 years in European ICUs: the VIP2 study. *Intensive Care Med*. 2020;46:57-69.
38. Brabrand M, Kellett J, Opio M, et al. Should impaired mobility on presentation be a vital sign? *Acta Anaesthesiol Scand*. 2018;62:945-952.
39. Hillman KM. Predicting outcome in the seriously ill-A new approach. *Acta Anaesthesiol Scand*. 2018;62:878-879.
40. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157-172; discussion 207-212.
41. Van Calster B, McLernon DJ, Van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17:230.
42. Perel P, Prieto-Merino D, Shakur H, et al. Predicting early death in patients with traumatic bleeding: development and validation of prognostic model. *BMJ*. 2012;345; e5166-e5166.
43. Shamout F, Zhu T, Clifton L, et al. Early warning score adjusted for age to predict the composite outcome of mortality, cardiac arrest or unplanned intensive care unit admission using observational vital-sign data: a multicentre development and validation. *BMJ Open*. 2019;9:1-7.
44. O'Neill SM, Clyne B, Bell M, et al. Why do healthcare professionals fail to escalate as per the early warning system (EWS) protocol? A qualitative evidence synthesis of the barriers and facilitators of escalation. *BMC Emerg Med*. 2021;21:15.
45. Kahneman D, Tversky A. On the psychology of prediction. *Psychol Rev*. 1973;80:237-251.