

Universiteit
Leiden
The Netherlands

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

Candel, B.G.J.

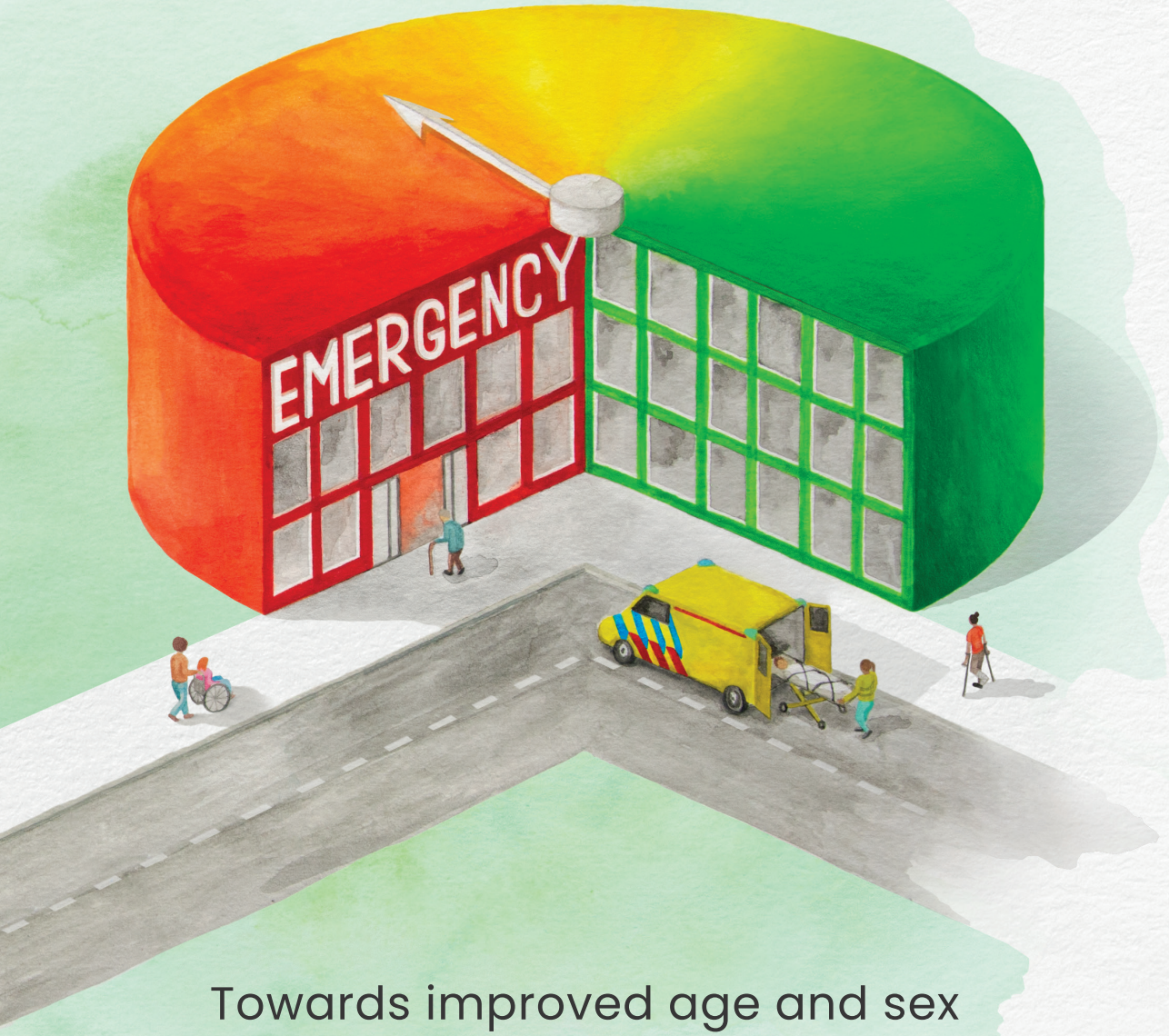
Citation

Candel, B. G. J. (2024, March 14). *Risk stratification in emergency medicine: towards improved age and sex adjusted risk assessment*. Retrieved from https://hdl.handle.net/1887/3721827

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3721827

Note: To cite this publication please use the final published version (if applicable).

Risk Stratification in Emergency Medicine



Towards improved age and sex adjusted risk assessment

B.G.J. Candel

Work hard in silence,
let your success be your noise.

Copyright ©
Auteur: **Bart Gerard Jan Candel**
Omslag: Meike Schipper
Provided by thesis specialist Ridderprint, ridderprint.nl
Printing: Ridderprint
Layout and design: Michèle Duquesnoy, persoonlijkproefschrift.nl

ISBN: 978-94-6483-408-6

Copyright 2023 © Bart Gerard Jan Candel
The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author.

Risk stratification in Emergency Medicine

Towards improved age and sex adjusted risk assessment

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 14 maart 2024
klokke 13.45 uur
door
Bart Gerard Jan Candel
geboren te Oosterhout NB
in 1990

Thesis committee

Promotor

Prof. Dr. E. de Jonge

Co-promotor

Dr B. de Groot (Radboud University Medical Center)

Other members

Prof. Dr. S.P. Mooijaart

Prof. Dr. R.H.H. Groenewold

Dr. S.M. Arbous

Dr. C.L. van den Brand (Erasmus Medical Center)

Prof. Dr. M. Brabrand (Odense University Hospital, Denmark)

The research described in this thesis was financially supported by the Dutch Emergency Medicine Research fund (SGO-fonds).

The publication of this thesis was financially supported by the Máxima Medical Center, the Leiden University Medical Center and the Dutch Society of Emergency Physicians (NVSHA).

TABLE OF CONTENTS

Chapter 1	General introduction	7
Part I	**Age and sex adjusted interpretation of physiological variables in Emergency Department patients**	13
Chapter 2	The association between vital signs and clinical outcomes in emergency department patients of different age categories	15
Chapter 3	Age-adjusted interpretation of biomarkers of renal function and homeostasis, inflammation, and circulation in Emergency Department patients	61
Chapter 4	The difference between the patients' initial and previously measured systolic blood pressure as predictor of mortality in older Emergency Department patients	87
Chapter 5	Sex differences in clinical presentation and risk stratification in the emergency department: an observational multicenter cohort study	105
Chapter 6	The association between systolic blood pressure and heart rate in Emergency Department patients: a multicentre cohort study	135
Part II	**Development and external validation of an age and sex adjusted Early Warning Score**	159
Chapter 7	The impact of age on predictive performance of National Early Warning Score at arrival to Emergency Departments: Development and External Validation	161
Chapter 8	Development and external validation of the International Early Warning Score for improved age and sex adjusted in-hospital mortality prediction in the Emergency Department.	187
Part III	**Bias in the use of risk tools to compare quality of care**	225
Chapter 9	The effect of treatment and clinical course during Emergency Department stay on severity scoring and predicted mortality risk in Intensive Care patients	227
Part IV	**Conclusions**	257
Chapter 10	General discussion	259
Chapter 11	Summary	267
Chapter 12	Dutch summary (Nederlandse samenvatting)	273
Appendices	References	282
	List of publications	295
	Curriculum vitae	297
	Dankwoord	298


EMERGENCY

1

GENERAL INTRODUCTION

1.1 AN AGING POPULATION

Emergency departments (EDs) are fast-paced working environments with a high turnover of acutely ill patients with varying disease presentations like fractures and wounds to sepsis and cardiac arrest. Over the last several decades, older people have represented an increasing proportion of ED visits worldwide due to an aging population.[1-3] The higher proportion of elderly represents a challenge for the acute care as older people more often present with non-specific complaints of an acute illness combined with multiple comorbidities, polypharmacy, and altered mental status, complicating diagnostic work-up and treatment.[2, 4, 5] More importantly, older patients are at higher risk for hospitalization and death after an ED visit.[6] Timely clinical decisions are therefore crucial. Many clinical decisions in the ED are based on the patients' risk and prognosis. Risk stratification is the process of assigning a risk status to patients, then using this information to direct care to improve overall health outcomes. As risk stratification enables clinicians to identify the right level of care and services for emergency patients, it may also help to prevent overcrowding through early disposition decisions. Risk stratification forms therefore the backbone of Emergency Medicine.

1.2 RISK STRATIFICATION

Until the 20th Century, risk stratification or prognostication was the major task of the medical profession before treatment options such as antibiotics existed. Doctors made the clinical diagnosis of pneumonia not because they were able to treat it, but only to prognosticate so they could inform the family to settle their affairs, as pneumonia used to be the main cause of death. After a rapid grow in medical developments, prognostication received almost no more attention in medical textbooks by 1970 due to a shift from prognostication to a diagnosis-based concept focused on diagnosis and therapy.[7] Explicit attention to prognosis consequently diminished and became an intrinsic part of diagnosis and therapy. Nowadays, in the day-to-day practice of medicine, clinicians are continuously making risk assessments by relying on their estimates of severity to make clinical decisions.[8] This is either done by intuitive, gestalt-based assessments or with the help of formal risk scores. In both ways, the clinician estimates the risk for adverse outcomes or deterioration based on a few or many physiological variables. Vital signs are the most important objective physiological variables used to assess the severity of illness,[9, 10] and abnormal vital signs are associated with poor outcomes.[11] However, age-related changes may affect the interpretation and use of vital signs during risk assessment. Thus, physiological variables may need different interpretations in older patients because of (patho)-

physiological changes with increasing age. For example, arterial walls become stiffer, causing higher blood pressure variability, and heart rate response to stress may be blunted.[12] Laboratory tests, or biomarkers, can also be used for risk assessment, but their normal values may also change with increasing age. For correct risk stratification, it is therefore important that clinicians understand the relationship between these physiological parameters and outcomes in both young and old patients.

1.3 SEX DIFFERENCES

Besides age differences in the ED, physicians have become more aware of sex differences in acute diseases, especially after the apparent sex differences in clinical outcomes during the Coronavirus Disease 2019 (COVID-19) pandemic.[13] Male patients had almost three times the odds of Intensive Care Unit (ICU) admission and 1.4 times higher odds of death with COVID-19.[13] Nonetheless, these outcome differences between men and women are not solely described for COVID-19. Sex differences in outcomes and clinical presentations have been described in many diseases, such as acute myocardial infarction, stroke, and infectious diseases.[14-20] Nonetheless, it remains unclear whether male sex is an independent risk factor for adverse outcomes in the overall ED population or if the higher risk in males is caused by higher disease severity at ED presentation. If healthcare providers are aware of sex differences in disease presentations and clinical outcomes, they may be able to better anticipate the clinical course and provide better treatment choices for both men and women. If male sex is a risk factor for poor outcomes in the ED, clinicians should incorporate this variable in their risk assessment.

1.4 RISK TOOLS

Many risk stratification tools have been introduced in the Emergency Department, such as Early Warning Scores (EWSs), the quick-sequential organ failure assessment score (qSOFA) for sepsis, the CURB-65 for pneumonia, and the Revised Trauma Score (RTS) for trauma patients.[21-26] Most scores detect clinical deterioration or predict mortality for all patients or subgroups. These scores were designed to support clinical intuition and gestalt by providing a formal risk score. Nonetheless, many risk scores have limited predictive performance in older patients.[27] This poor performance may be caused by relevant age differences in risk which are often not considered while developing these scores. Moreover, many scores use single thresholds for physiological variables (i.e., higher, or lower than a systolic blood pressure of 90mmHg), suggesting a clinically relevant change

in prognosis at this threshold. Also, using one threshold indicates that one cut-off may discriminate between a good and bad prognosis. However, recent studies in patients with traumatic brain injury and sepsis showed that prognosis linearly worsens with decreasing systolic blood pressure (SBP), without a single identifiable threshold.[28, 29] Similarly, linear, or U-shaped instead of dichotomous associations may exist for other vital signs used in ED risk stratification. Although most EWSs use multiple thresholds for each physiological variable, they were not developed by applying statistical methods but rather by clinical consensus.[10, 30] Using scores that are not age and sex-adjusted with erroneously selected thresholds may negatively affect patient care if used for clinical guidance. This may especially apply to the ED where patients are often undifferentiated and at risk for deterioration. An example of an excellent clinical score is the Acute Physiology and Chronic Health Evaluation 4th edition (APACHE-IV) used for Intensive Care Unit (ICU) patients.[31, 32] The APACHE-IV has excellent discriminative performance and calibration for ICU patients, not only allowing to predict accurately the outcome for an individual patient, but also allowing to compare the quality of care of ICUs by calculating a Standardized Mortality Ratio (SMR). The SMR is the observed mortality in ICU patients divided by the predicted mortality. An SMR below 1 implies fewer people died after ICU admission than predicted, suggesting good quality of care. However, although this score is an excellent prediction tool, other factors may lead to a higher calculated SMR. For example, physiological parameters may improve or deteriorate in the hours before ICU admission,[33-36] e.g., during initial treatment in the emergency department (ED). These data measured before ICU admission are not included in the original APACHE-IV score.[31, 32] Consequently, the calculated severity of illness and predicted mortality risk may differ when the same initial treatment is provided before or after ICU admission for a given patient.[37, 38] It is essential to understand the effect of ED treatment and clinical course on the predicted mortality risk calculated by the APACHE-IV, if these scores are used to compare the quality of care among departments or hospitals.

1.5 THE NETHERLANDS EMERGENCY DEPARTMENT EVALUATION DATABASE (NEED)

This thesis used the Netherlands Emergency department Evaluation Database (NEED) to investigate the aims of this thesis. The NEED is the national quality registry for EDs in the Netherlands, founded in 2016, and aims to contribute to the transparency and quality of ED care in the Netherlands by providing reliable data to the participating EDs (www.stichting-need.nl). Participating EDs receive twice a year a report for

benchmarking in which the data of the hospital are compared with the pooled data of all participating EDs in the NEED.

1.6 AIM OF THE THESIS

This thesis has four aims. First, to investigate whether the interpretation of physiological variables should be age and sex adjusted for ED patients. The second aim is to assess whether thresholds for vital signs and biomarkers truly exist. The third aim is to develop and externally validate an age and sex adjusted Early Warning Score and assess whether this score improves the classification of ED patients into low, intermediate, and high risk. The fourth aim is to evaluate the effect of initial ED treatment versus immediate ICU admission on predicted mortality risk and SMR in critically ill patients.

1.7 OUTLINE OF THE THESIS

This thesis is divided into three parts. The first part includes the first and second aim of this thesis. In **chapter 2,** we study the association between vital signs and clinical outcomes in ED patients of different age categories. In **chapter 3,** we evaluate the effect of age on the interpretation of biomarkers of renal function and homeostasis, inflammation, and circulation in ED patients. In **chapter 4,** we investigate whether the difference between the patients' initial and previously measured systolic blood pressure (delta SBP) predicts mortality in older ED patients. In **Chapter 5,** we study sex differences in clinical presentation and risk stratification in the ED. In **Chapter 6,** the association between SBP and heart rate (HR) is investigated.

The second part of this thesis is devoted to developing an age and sex adjusted EWS. **Chapter 7** describes the effect of including age in the National Early Warning Score (NEWS) on its predictive performance in different age categories. In **Chapter 8,** we develop and externally validate an age and sex-adjusted International Early Warning Score (IEWS).

The third part of this thesis includes one chapter, **Chapter 9**, in which we study the effect of treatment and clinical course during ED stay on severity scoring and predicted mortality risk in ICU patients.

Finally, in **chapter 10,** the main conclusions of this thesis are summarized and discussed, and future perspectives are proposed.



PART

I

AGE AND SEX ADJUSTED INTERPRETATION OF PHYSIOLOGICAL VARIABLES IN EMERGENCY DEPARTMENT PATIENTS.


EMERGENCY

2

THE ASSOCIATION BETWEEN VITAL SIGNS AND CLINICAL OUTCOMES IN EMERGENCY DEPARTMENT PATIENTS OF DIFFERENT AGE CATEGORIES.

Bart GJ Candel
Renée Duijzer
Menno I Gaakeer
Ewoud ter Avest
Ozcan Sir
Heleen Lameijer
Roger APA Hessels
Resi Reijnen
Erik W van Zwet
Evert de Jonge
Bas de Groot

Published in the Emergency Medicine Journal, 2022 – Volume 39, issue 12, p903-911

ABSTRACT

Background Appropriate interpretation of vital signs is essential for risk stratification in the Emergency Department (ED) but may change with advancing age. In several guidelines, risk scores such as SIRS and qSOFA, commonly used in emergency medicine practice (as well as critical care) specify a single cut-off or threshold for each of the commonly measured vital signs. Although a single cut-off may be convenient, it is unknown whether a single cut-off for vital signs truly exists and if the association between vital signs and in-hospital mortality differs per age-category.

Aims To assess the association between initial vital signs and case-mix adjusted in-hospital mortality in different age categories.

Methods Observational multi-centre cohort study using the Netherlands Emergency department Evaluation Database (NEED) in which consecutive ED patients ≥18 years were included between 1 January 2017 and 12 January 2020. The association between vital signs and case-mix adjusted mortality were assessed in three age categories (18-65; 66-80; >80 years) using multivariable logistic regression. Vital signs were each divided in five to six categories, e.g., systolic blood pressure (SBP) categories (≤80, 81–100, 101–120, 121-140, >140mmHg).

Results We included 101,416 patients of whom 2374 (2.3%) died. Adjusted Odds Ratios for mortality increased gradually with decreasing SBP and decreasing peripheral oxygen saturation (SpO_2). Diastolic blood pressure (DBP), mean arterial pressure (MAP) and heart rate (HR) had quasi-U-shaped associations with mortality. Mortality did not increase for temperatures anywhere in the range between 35.5-42.0°C, with a single cut-off around 35.5 °C below which mortality increased. Single cut-offs were also found for MAP <70mmHg and respiratory rate >22/min.. For all vital signs, older patients had larger increases in absolute mortality compared to younger patients.

Conclusion For SBP, DBP, SpO_2, and HR no single cut-off existed. The impact of changing vital sign categories on prognosis was larger in older patients. Our results have implications for the interpretation of vital signs in existing risk stratification tools and acute care guidelines.

INTRODUCTION

Vital signs are used for risk stratification in nearly all acute care guidelines.[21, 39-41] Many risk stratification tools, such as SIRS, qSOFA, CURB-65,[21-25] specify a single cut-off (or threshold) for each of the commonly measured vital signs, suggesting that one cut-off may discriminate between good and bad prognosis. However, recent studies in patients with traumatic brain injury and sepsis showed that prognosis linearly worsened with decreasing systolic blood pressure (SBP), without an identifiable single threshold.[28, 29] Similarly, linear, or U-shaped instead of dichotomous associations may exist for other vital signs used in ED risk stratification.

In addition, most risk stratification models are applied in all ED patients irrespective of age. However, a recent study showed that risk stratification tools are inappropriate for older ED patients.[27, 42] The association between vital signs and outcomes may change with advancing age due to physiological changes. Arterial walls become stiffer in older patients, heart rate response to stress is blunted and fever is absent in many older patients with sepsis.[12]

Although a single cut-off for vital signs may be convenient in risk stratification tools, numerical scores based on predicted mortality will possibly better inform physicians and prevent impaired recognition of early deterioration in ED patients. Physiological changes with increasing age may implicate that currently used risk stratification tools should be age-adjusted. This is particularly important given the ageing of the population with an increasing number of older ED patients.

Therefore, the aims of the present study are two-fold: First, to assess the associations of vital signs and relevant clinical outcomes (e.g., mortality and Intensive Care Unit (ICU)/Medium Care Unit (MCU) admission) and whether a single cut-off or threshold exists for each vital sign, and second, to study whether associations of vital signs and relevant clinical outcomes change with advancing age.

METHODS

Study design and setting

This observational multi-centre study was conducted in three EDs in the Netherlands, each with approximately 25,000-30,000 ED visits per year. Data from the three sites spanned slightly different times: data for the tertiary centre included visits between 1 January 2017 – 8 June 2019, and data from the two urban teaching hospitals were from 1 January 2019 – 12 January

2020 1 January 2017 – 31 December 2019, respectively. The tertiary care centre uses the Manchester Triage System (MTS) and both urban hospitals use the Dutch Triage Standard (NTS). The study was approved by the medical ethics committee of the LUMC. The study was registered in the Netherlands Trials Register: NL8422.

Selection of participants
All consecutive ED patients ≥18 years were included if one or more of the following vital signs were measured: respiratory rate (RR), peripheral oxygen saturation (SpO_2), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial blood pressure (MAP), heart rate (HR) and temperature.

Data collection
Data were collected from the Netherlands Emergency department Evaluation Database (NEED), the Dutch quality registry for EDs (www.stichting-need.nl), a record containing clinical data from all ED visits from the participating hospitals. Supplementary file 1 provides a detailed description of the data collected in the NEED. Twice a year all clinical data from the participating hospitals is downloaded into the record. Onset of participation in the NEED differs per hospital. For the current study, data were available from three of the nine participating hospitals. The NEED stores data using the web-based application Project Manager Internet Server (ProMISe, Leiden, the Netherlands, https://www.msbi.nl/promise/promise.aspx). Privacy sensitive data are encrypted by a trusted third party (ZorgTTP, Houten, the Netherlands, www.zorgttp.nl) with Trusted Reversible Encryption Service.

The vital signs measured at the beginning of ED presentation, before ED treatment, were recorded in the NEED. Only one set of vital signs was recorded per patient. If patients arrived by own transport, HR, SpO_2 and temperature were measured in a triage room by a nurse mostly in seated, sometimes supine, position. These vital signs were entered manually in the hospital information system. Once moved to a patient room, the other vital signs were measured by a nurse with the MP52 IntelliVue, Intellivue MP30, Intellivue MX400 or MX500 monitor (Philips, Eindhoven, the Netherlands). Multi-patient reusable latex-free NIBP Comfort Care Cuff from Philips were used with cuff a small cuff (27-35cm), medium cuff (35-45cm), large cuff (42-54cm) and extra-large cuff (44-56cm), depending on the size of the patient. Vital signs registered in the patient room were automatically transferred in the hospital information system, except for temperature, which was entered manually. Temperature was measured with a Genius™ 3 Tympanic Thermometer. If patients arrived by ambulance, vital signs were measured in supine position in the patient room as described. In the tertiary care centre,

vital signs were automatically measured every 2.5 to 15 min and transferred to the electronic patient file with PDMS software (Chipsoft, the Netherlands), depending on the personalized setting for a specific patient. Median values of vital signs were registered for the first 10 minutes of measuring.

Measurements
Before we could perform statistical analysis, several variables had to be synchronized among hospitals. Beforehand we could not assume that the association between vital signs and outcomes were linear. Therefore, vital signs were categorized into ranges based on expected distribution, and commonly used reference intervals: RR (Not measured, 0-9, 10-19, 20-29, ≥30/min), SpO_2 (0-80, 81-85, 86-90, 91- 95, 96-100%), SBP (0-80, 81-100, 101-120, 121-140, >140mmHg), DBP (0-60, 61-80, 81-100, 101-120, >120mmHg), MAP (0-60, 61-80, 81-100, 101-120, >120mmHg), HR (0-50, 51-75, 76-100, 101-125, >125beats/min.), temperature (Not measured, 0-30, 31-34, 35-37, 38-39, ≥40°C).

Correspondent to literature,[43] RR and temperature were often not registered. To limit the number of missings in the analyses, we included a category "not registered". In EDs in the Netherlands, vital signs are not registered for all patients. For example, vital signs are often not registered in patients with an ankle distortion or a single fracture, because they are at low risk of adverse events and often discharged, and RR is often only registered if patients are considered critically ill by the nurse. Therefore, the fact that vital signs are registered may provide relevant information for risk stratification. Supplementary file 2 describes how the variables triage levels, chief complaints (the top ten chief complaints was used) and treating specialties were modified before they could be used as potential confounders in the statistical analyses.

Patients were stratified into three age categories: 18-65, 66-80 and >80 years. These age categories were chosen based on the mean SBP in the general population in the Netherlands according to the National Institute for Public Health and the Environment (RIVM, www.rivm.nl). The mean SBP increases substantially above 65 years.

Outcome
The primary outcome was whether there was a vital sign category that could be used as a cut-off to predict the outcomes of in-hospital mortality and ICU or MCU admission, which also included the Coronary Care Unit.

Sample size estimation
Approximately five to ten events per variable are needed to prevent overfitting in association studies.[44] The NEED contained 148,828 ED visits of patients ≥18 years. We estimated that in ~60% of the ED visits vital signs were registered resulting in ~90,000 ED visits which could be used for the analyses. In-hospital mortality was estimated to be ~3%. Included patients were stratified in three age categories, yielding ~90,000 / 3 = 30,000 patients per age category. Thus, we expected to have ~0.03 x 30,000 = 900 events per group to be able to adjust for the 37 potential confounders (see main statistical analysis) in the analyses with in-hospital mortality as outcome, which is appropriate to prevent overfitting.

Statistical analyses
Patient characteristics were described as mean (standard deviation (SD)) for normally distributed data, and median (interquartile range (IQR)) for skewed data. Categorical data were presented as number (N, (%)).

For all analyses, one category per vital sign was used as reference category (i.e. reference category for HR was 50-75 beats/min.). The number of measured vital signs outside the reference range were calculated.

We used multivariable logistic regression analyses for the association between vital sign categories and the primary and secondary outcomes. The following potential confounders were entered in the models for which then backward stepwise elimination was performed: age, sex, triage level, top ten chief complaints, hospital, treating specialty, disposition (ward, ICU/MCU), SpO_2, temperature (including category 'not registered'), GCS (not assessed, assessed and 15 points, assessed and <15 points), lab tests (0=no lab tests performed, 1= any lab test performed), blood gas analyses (0= not tested, 1= tested), number of consultations in the ED (0, 1, 2 or ≥3 consultations), and performed radiological test (0= none, 1= any test). Other vital signs were not entered in the models because of potential multicollinearity among vital signs. The number of consultations, lab tests and radiological tests were used as measures of comorbidities/complexity, as has been suggested previously.[45, 46] Triage level, GCS, blood gas analyses and ICU/MCU admission were used as disease severity measures.[27] ICU/MCU admission is not only an outcome, but also a proxy of disease severity not captured in the initial vital signs. For example, a patient with an initial SBP of 90 mmHg who responds to fluid resuscitation can be admitted to a normal ward and has a different risk for in-hospital mortality compared to a patient who also has an initial SBP of 90 mmHg does not respond to fluid resuscitation and consequently needs ICU admission. In other words, the variable ICU/

MCU admission also reflects the response to treatment. For the secondary outcome, the potential confounder disposition was replaced by the amount of fluid administration as a proxy of disease severity. Multicollinearity was assumed not to be a problem if variation inflation factors (VIF) were below three. The associations between vital sign categories and case-mix adjusted outcomes were assessed in two ways to give better insight in differences between age categories, as the relative risk for the outcome may be similar among age categories, while the absolute risk may be different due to different baseline risk at the reference values. We used adjusted odds ratios (AORs) with 95%-confidence intervals (95%-CI) as relative risk measure and predicted (case mix-adjusted) mortality as absolute risk measure.

In the figures, the AORs of in-hospital mortality were plotted as a function of vital sign categories for the pooled data and for all age categories. If we observed a sudden rise or decline in AORs by changing vital sign categories, we performed minimal p-value analyses to assess a more precise value for the possible threshold. In these analyses, we changed the vital sign into a binary value (below or above a cut-off value) in the multivariable logistic regression as described above. We repeated the analyses with all possible cut-off values in steps of 0.5°C (temperature), 1/min. (RR), 5.0mmHg (SBP, DBP, MAP), 5.0 beats/min. (HR) or 5.0% (SpO_2). The cut-off value with the lowest p-value for the AOR was considered as threshold. In this way, a threshold was determined in approximately continuous data of the vital sign. Thresholds with minimal p-value analyses were not determined if vital signs had no clear single cut-off.

A P-value<0.05 was considered as statistically significant. Data were analysed using SPSS (version 25.0, IBM, New York, USA).

Subgroup and sensitivity analyses

We performed subgroup analyses for suspected infection and trauma patients to assess whether aetiology affected the primary associations of interest. See supplementary files 3 and 4 for case definitions. A sensitivity analyses was performed to assess whether hospital transfers may have influenced our results, as we did not follow-up on these patients. The main analyses were repeated in which patients transferred to other hospitals were considered as deceased.

RESULTS

Patient inclusion and characteristics

Figure 1 shows that of the 147,728 ED visits of patients ≥18 years, 101,416 patients were included in whom vital signs were measured.

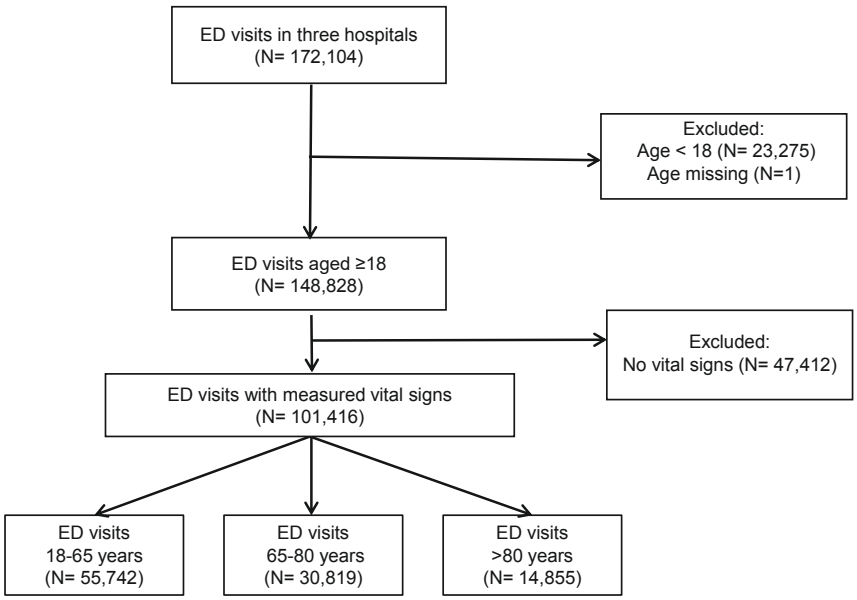


Figure 1. Patient inclusion and flow through the study. Observational multi-centre cohort study using the Netherlands Emergency department Evaluation Database (NEED). Three different age categories were included in which vital signs were measured.

Table 1 shows patient characteristics of the total cohort and per age-category. The mean age of included patients was 59.6 years (19.6). Table 1 also shows the percentage of patients in whom vital signs fell outside the reference ranges which are normally used in adults. For most values, the rate of abnormal vital signs increased with advancing age.

Table 1. Patient characteristics for the total cohort and different age categories.

	Total cohort N= 101,416 (100%)	18-65 N=55,742 (55%)	66-80 N=30,819 (30%)	>80 N=14,855 (15%)
Demographics N, (%)				
Sex, female	50,333 (49.6)	28,226 (50.7)	13,752 (44.6)	8315 (56.0)
Age, mean (SD)	59.6 (19.6)	45.3 (14.1)	72.9 (4.2)	85.8 (3.9)
Hospital setting N, (%)				
LUMC	28,361 (28.0)	17,980 (32.3)	7536 (24.5)	2845 (19.2)
MCL	53,378 (52.6)	27,048 (48.5)	17,698 (57.4)	8632 (58.1)
CZE	19,677 (19.4)	10,714 (19.2)	5585 (18.1)	3378 (22.7)
Treating specialty N, (%)				
Emergency medicine	17,656 (17.4)	11,062 (19.8)	4274 (13.9)	2320 (15.6)
Surgical	19,434 (19.2)	11,902 (21.4)	5076 (16.5)	2456 (16.5)
Medical	62,262 (61.4)	31,096 (55.8)	21,157 (68.6)	10,009 (67.4)
Top-10 chief complaints N, (%)				
Feeling unwell	20,414 (20.1)	9192 (16.5)	7392 (24.4)	3830 (26.3)
Abdominal pain	14,185 (14.0)	10,136 (18.2)	3035 (10.0)	1014 (7.0)
Dyspnoea	11,937 (11.8)	4975 (8.9)	4673 (15.4)	2289 (15.7)
Chest pain	11,067 (10.9)	6327 (11.4)	3523 (11.6)	1217 (8.4)
Extremity problems	8106 (8.0)	3736 (6.7)	2422 (8.0)	1948 (13.4)
Collapse	3868 (3.8)	1722 (3.1)	1433 (4.7)	713 (4.9)
Trauma	3778 (3.7)	2263 (4.1)	876 (2.9)	639 (4.4)

Table 1. Patient characteristics for the total cohort and different age categories. (continued)

	Total cohort N= 101,416 (100%)	**18-65 N=55,742 (55%)**	**66-80 N=30,819 (30%)**	**>80 N=14,855 (15%)**
Palpitations	3376 (3.3)	1665 (3.0)	1370 (4.5)	341 (2.3)
Wounds	2256 (2.2)	1442 (2.6)	601 (2.0)	213 (1.5)
Headache	2144 (2.1)	1465 (2.6)	475 (1.6)	204 (1.4)
Miscellaneous	18,311 (18.2)	11,437 (22.9)	4465 (14.9)	2147 (14.7)
Vital signs Mean (SD) [N] {% outside of reference category*}				
RR (/min.)	18 (6) [68,662] {23.4}	17 (6) [34,923] {18.3}	19 (6) [22,545] {27.1}	19 (6) [11,194] {32.1}
SPO2 (%)	97 (6) [91,785] {14.2}	99 (3) [49,628] {8.5}	96 (9) [28.374] {18.1}	96 (4) [13,783] {23.6}
SBP (mmHg)	133 (34) [89,531] {79.1}	128 (28) [47,444] {76.3}	137(40) [28,189] {81.7}	141 (34) [13,898] {83.3}
DBP (mmHg)	83 (17) [89,475] {60.7}	84(16) [47,424] {62.5}	81(18) [28.169] {58.7}	80 (19) [13,882] {58.3}
MAP (mmHg)	99(19) [89,512] {50.9}	98(18) [47,438] {47.9}	100 (21) [28.181] {53.8}	100 (21) [12,407] {54.7}
HR (beats/min.)	86 (52) [84,776] {23.4}	87(69) [44,611] {18.3}	86 (22) [26978] {27.1}	83 (21) [13,187] {32.1}
Temperature (°C)	37.0 (2.0) [82,364] {42.1}	37.0 (0.8) [45,073] {39.7}	37.0 (3.3) [25.349] {44.6}	36.9 (0.9) [11,942] {45.8}

Table 1. Patient characteristics for the total cohort and different age categories. (continued)

	Total cohort N= 101,416 (100%)	**18-65 N=55,742 (55%)**	**66-80 N=30,819 (30%)**	**>80 N=14,855 (15%)**
Proxies of disease severity N, (%)				
Triage level				
Blue and Green	22,414 (22.1)	12,952 (23.2)	6222 (20.2)	3240 (21.8)
Yellow	43,804 (43.2)	24,511 (44)	13,021 (42.2)	6272 (42.2)
Orange	28,176 (27.8)	14,750 (26.5)	9198 (29.8)	4228 (28.5)
Red	5339 (5.3)	2590 (4.6)	1896 (6.2)	853 (5.7)
GCS				
Not assessed	91,143 (89.9)	51,021 (91.5)	27,288 (88.5)	12,834 (86.4)
15	9041 (8.9)	4154 (7.5)	3170 (10.3)	1717 (11.6)
<15	1232 (1.2)	567 (1.0)	361 (1.2)	304 (2)
Fluid administration				
None	82,182 (81)	45,626 (81.9)	24,623 (79.9)	11,933 (80.3)
0-500ml	9164 (9.0)	4734 (8.5)	2909 (9.4)	1521 (10.2)
>500ml	10,070 (9.9)	5382 (9.7)	3287 (10.7)	1401 (9.4)
Blood gas analyses	21,750 (21.4)	9874 (17.7)	7924 (25.7)	3952 (26.6)

Table 1. Patient characteristics for the total cohort and different age categories. (continued)

	Total cohort N= 101,416 (100%)	**18-65 N=55,742 (55%)**	**66-80 N=30,819 (30%)**	**>80 N=14,855 (15%)**
Proxies of comorbidities/complexity N, (%)				
Number of consultations				
None	34,657 (34.2)	20,575 (36.9)	9671 (31.4)	4411 (29.7)
1	57,810 (57.0)	30,107 (54)	18,633 (60.5)	9070 (61.1)
2	6369 (6.3)	3193 (5.7)	2012 (6.5)	1164 (7.8)
>2	829 (0.8)	417 (0.7)	257 (0.8)	155 (1.0)
≥1 lab test	84,434 (83.3)	44,232 (79.4)	27,144 (88.1)	13,058 (87.9)
Radiological test**	60,739 (59.9)	29,741 (53.4)	19,908 (64.6)	11,090 (74.7)

Abbreviations: N= number, SD = standard deviation, LUMC= Leiden University Medical Centre, MCL= Medical Centre Leeuwarden, CZE= Catharina Hospital Eindhoven, RR = respiratory rate, SPO2 = peripheral oxygen saturation, SBP = systolic blood pressure, DBP = diastolic blood pressure, MAP = mean arterial pressure, HR = heartrate, GCS = Glasgow Coma Scale, mmHg = millimetre of mercury, °C= degrees Centigrade, ml = millilitres.

*Used reference categories were: SBP 100-140mmHg, DBP 60-90mmHg, MAP 70-100mmHg, HR 60-80beats/min., RR 12-20/min. SpO_2 ≥95%, temperature 36.5-37.5°C.

** If one or more of the following radiological tests were performed: Ultrasound, radiography, and computer-tomography.

Table 2. Adjusted Odds Ratios (AORs) for the association between vital signs and in-hospital mortality by different age groups.

	Total cohort	18-65 years	66-80 years	>80 years
Systolic blood pressure (mmHg)				
>140*	1.00			
121-140	1.35 (1.18-1.54)	1.68 (1.22-2.33)	1.28 (1.04-1.57)	1.39 (1.12-1.72)
101-120	1.91 (1.67-2.19)	2.50 (1.81-3.45)	1.75 (1.42-2.16)	1.97 (1.57-2.46)
81-100	2.62 (2.22-3.09)	2.82 (1.92-4.13)	2.83 (2.21-3.61)	2.44 (1.84-3.25)
0-80	4.07 (3.31-5.01)	6.38 (4.20-9.70)	3.33 (2.41-4.60)	4.17 (2.87-6.05)
Diastolic blood pressure (mmHg)				
>120	1.73 (1.37-2.19)	1,82 (1.08-3.07)	2.25 (1.57-3.20)	1.26 (0.84-1.88)
101-120	1.15 (0.96-1.37)	1.48 (1.04-2.10)	1,24 (0.95-1.63)	0.88 (0.65-1.20)
81-100*	1.00			
61-80	1.23 (1.09-1.39)	1.62 (1.24-2.11)	1,30 (1.08-1.56)	0.99 (0.82-1.21)
0-60	2.12 (1.84-2.44)	3.42 (2.47-4.73)	2,32 (1.87-2.88)	1.59 (1.27-2.00)
Mean Arterial blood pressure (mmHg)				
>120*				
101-120	0.88 (0.75-1.03)	1.01 (0.68-1.50)	0.81 (0.64-1.02)	0.95 (0.74-1.23)
81-100	1.06 (0.90-1.23)	1.56 (1.07-2.27)	0.88 (0.70-1.12)	1.11 (0.86-1.42)
61-80	1.73 (1.47-2.05)	2.02 (1.33-3.07)	1.72 (1.33-2.21)	1.75 (1.32-2.30)
0-60	3.27 (2.60-4.10)	5.72 (3.48-9.42)	2.86 (2.01-4.09)	3.17 (2.17-4.64)

Table 2. Adjusted Odds Ratios (AORs) for the association between vital signs and in-hospital mortality by different age groups. (continued)

	Total cohort	18-65 years	66-80 years	>80 years
Peripheral oxygen saturation (%)				
96-100*	1.00			
91-95	1.46 (1.31-1.62)	1.68 (1.30-2.18)	1.30 (1.10-1.53)	1.48 (1.25-1.77)
86-90	2.18 (1.83-2.61)	3.28 (2.17-4.94)	1.84 (1.41-2.39)	2.22 (1.66-2.97)
81-85	3.58 (2.80-4.58)	3.70 (2.08-6.57)	2.66 (1.81-3.92)	4.92 (3.32-7.29)
0-80	6.28 (5.05-7.82)	11.1 (7.40-16.6)	4.50 (3.19-6.34)	4.99 (3.30-7.56)
Respiratory rate (/min)				
Not registered*	1.00			
0-9	0.99 (0.49-2.01)	0.79 (0.24-2.69)	1.54 (0.58-4.07)	0.58 (0.08-4.46)
10-19	1.10 (0.92-1.30)	0.82 (0.60-1.13)	1.15 (0.87-1.51)	1.41 (1.02-1.94)
20-29	1.83 (1.53-2.18)	1.47 (1.04-2.06)	1.89 (1.43-2.50)	2.28 (1.64-3.15)
>30	2.89 (2.32-3.60)	2.57 (1.62-4.06)	3.23 (2.31-4.53)	3.17 (2.15-4.68)
Heart rate (beats/min)				
>125	2.98 (2.44-3.64)	4.17 (2.77-6.29)	2.47 (1.79-3.41)	2.76 (1.95-3.90)
101-125	2.50 (2.16-2.91)	2.61 (1.85-3.69)	2.71 (2.16-3.40)	2.22 (1.72-2.86)
76-100	1.43 (1.25-1.63)	1.34 (0.97-1.85)	1.48 (1.21-1.82)	1.48 (1.21-1.81)
51-75*				
0-50	1.61 (1.19-2.17)	2.04 (1.02-4.08)	1.67(1.04-2.69)	1.54 (0.96-2.45)

Table 2. Adjusted Odds Ratios (AORs) for the association between vital signs and in-hospital mortality by different age groups. (continued)

	Total cohort	18-65 years	66-80 years	>80 years
Temperature (°C)				
Not registered*				
≥40	0.49 (0.28-0.86)	0.40 (0.12-1.33)	0.43 (0.17-1.10)	0.65 (0.26-1.60)
38-39	0.68 (0.57-0.80)	0.69 (0.47-1.00)	0.74 (0.57-0.95)	0.57 (0.42-0.78)
35-37	0.89 (0.78-1.00)	0.91 (0.69-1.18)	0.81 (0.67-0.98)	0.97 (0.78-1.20)
31-34	2.74 (1.96-3.82)	4.11 (2.25-7.50)	2.10 (1.23-3.58)	2.77 (1.52-5.04)
0-30	1.54 (0.52-4.54)	1.17 (0.15-9.34)	0.60 (0.06-5.74)	4.04 (0.64-25.6)

*= reference category in the regression analyses
The following potential confounders were entered in the model for through backward stepwise regression: age, gender, top ten chief complaints, hospital, treating specialty, SpO_2, temperature, proxies of disease severity (triage level, Glasgow Coma Scale, blood gas analyses, Intensive Care Unit admission) and proxies of disease comorbidities/complexity (number of consultations, lab tests, performed radiological tests).

The association between vital signs and clinical outcomes

The AORs for predicted mortality (table 2) increased gradually with worsening values of SBP and SpO_2, while DBP, MAP and HR had a quasi-U-shaped association with mortality (Figures 2 and 3). There was no clear cut-off point between categories of SBP, DBP, SpO_2 and HR and AORs for mortality. The AOR for RR gradually increased between 10-19mm (Figure 3), with a substantial increase for mortality at 22/min. The AOR of mortality showed a substantial increase for temperature below a threshold of 35.5°C; however, temperatures above this were not associated with mortality. For MAP, a threshold at <70mmHg was found for mortality. Although single cut-offs for MAP, RR and temperature exist, using these cut-offs would ignore further increase of risk with more extreme values of these vital signs.

Older patients had larger increases in case-mix adjusted predicted mortality compared to younger people with changing vital sign categories (right panels in figure 2 and 3). In contrast, AORs for mortality decreased or increased to a similar extent with changing vital sign categories. Only patients aged 18-65 years with an SpO_2 <80% had higher odds for mortality (AOR 11.09; 7.40-16.62) compared to patients aged 66-80 years (AOR 4.50; 3.19-6.34) and >80 years (AOR 4.99; 3.30-7.56).

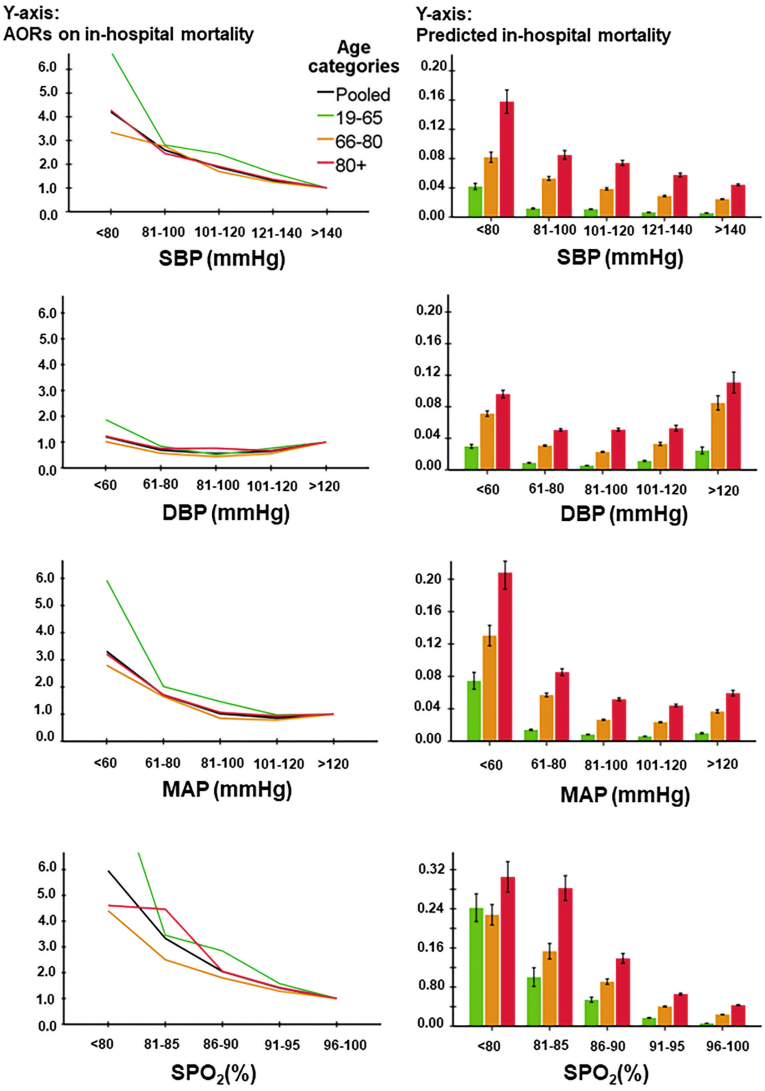


Figure 2. Adjusted odds ratios (AORs) for systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP) and peripheral oxygen saturation (SpO_2) for in-hospital mortality (left side) and predicted in-hospital mortality (right side) as a function of vital signs in different age categories. Note that the black line is the pooled data of all three age categories together. The following potential confounders were entered in the model for in-hospital mortality through backward stepwise regression: age, gender, triage level, top ten chief complaints, hospital, treating specialty, disposition, SpO_2 (except for association of SpO_2), temperature (except for association of temperature), Glasgow coma scale, lab tests, number of consultations in the emergency department and performed radiological test.

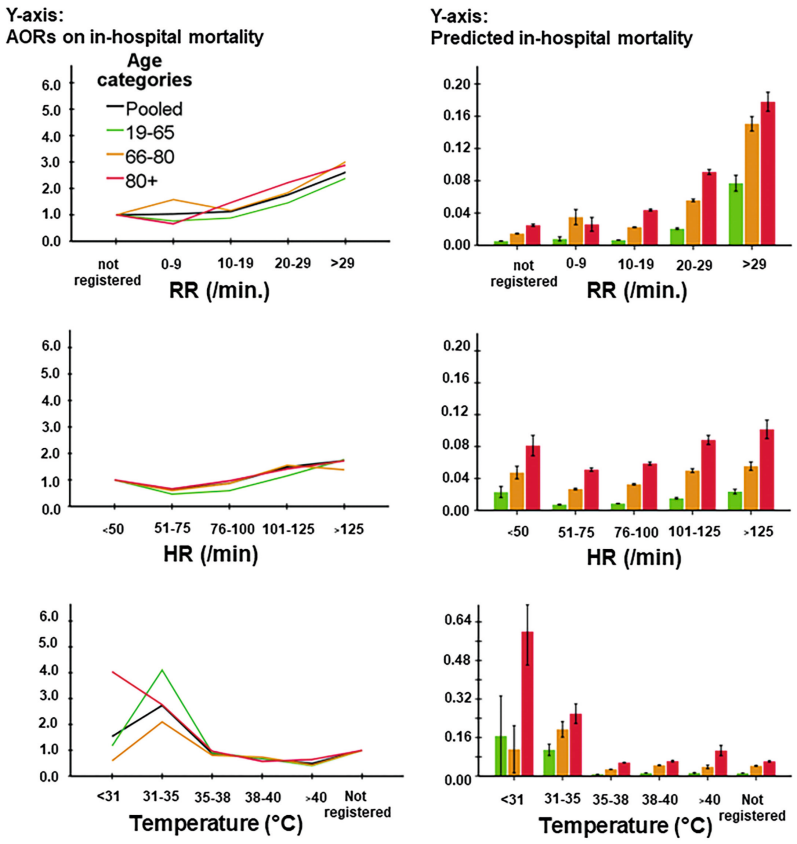


Figure 3. Adjusted odds ratios (AORs) for respiratory rate (RR), heart rate (HR) and temperature for in-hospital mortality (left side) and predicted in-hospital mortality (right side) as a function of vital signs in different age categories. Note that the black line is the pooled data of all three age categories together. The following potential confounders were entered in the model for in-hospital mortality through backward stepwise regression: age, gender, triage level, top ten chief complaints, hospital, treating specialty, disposition, SpO_2, temperature, Glasgow coma scale, lab tests, number of consultations in the emergency department and performed radiological test.

For ICU/MCU admission (table 3, supplementary files 5 and 6), SBP had a relevant cut-off in the category of <80mmHg and was calculated at 70mmHg with minimal p-value analysis. For MAP, a threshold of <60mmHg existed for admission to the ICU/MCU. When RR was registered, AOR for ICU admission was at least doubled compared to when it was not recorded, but no other cut-off was seen. Temperature < 31 °C had a substantially higher likelihood for ICU admission.

Table 3. Adjusted Odds Ratios (AORs) for the association between vital signs and ICU/MCU admission by different age groups.

	Total cohort	18-65 years	66-80 years	>80 years
Systolic blood pressure (mmHg)				
>140*	1.00			
121-140	0.89 (0.82-0.97)	0.90 (0.79-1.03)	0.95 (0.83-1.09)	0.94 (0.75-1.17)
101-120	0.98 (0.89-1.08)	1.01 (0.88-1.17)	0.95 (0.81-1.12)	1.14 (0.88-1.46)
81-100	0.97 (0.84-1.11)	0.97 (0.79-1.18)	0.93 (0.74-1.16)	1.21 (0.86-1.72)
0-80	1.94 (1.64-2.30)	2.11 (1.63-2.72)	1.98 (1.50-2.60)	1.88 (1.21-2.94)
Diastolic blood pressure (mmHg)				
>120	2.18 (1.85-2.58)	2.34 (1.83-3.00)	2.31 (1.76-3.03)	1.54 (0.98-2.40)
101-120	1.34 (1.21-1.48)	1.30 (1.12-1.51)	1.36 (1.15-1.61)	1.34 (1.01-1.77)
81-100*	1.00			
61-80	0.93 (0.86-1.00)	0.99 (0.88-1.12)	0.82 (0.72-0.93)	1.00 (0.81-1.22)
0-60	1.32 (1.18-1.49)	1.71 (1.41-2.07)	1.32 (1.10-1.58)	1.04 (0.78-1.37)
Mean Arterial blood pressure (mmHg)				
>120*	1.00			
101-120	0.74 (0.67-0.82)	0.69(0.60-0.81)	0.81 (0.70-0.95)	0.73 (0.56-0.93)
81-100	0.64 (0.57-0.70)	0.65(0.56-0.76)	0.62 (0.52-0.73)	0.72 (0.56-0.93)
61-80	0.76 (0.67-0.86)	0.78(0.64-0.95)	0.80 (0.66-0.98)	0.73 (0.53-1.00)
0-60	1.70 (1.40-2.06)	2.18(1.61-2.96)	1.89 (1.38-2.58)	1.28 (0.81-2.02)

Table 3. Adjusted Odds Ratios (AORs) for the association between vital signs and ICU/MCU admission by different age groups. (continued)

	Total cohort	18-65 years	66-80 years	>80 years
Peripheral oxygen saturation (%)				
96-100*	1.00			
91-95	1.16 (1.06-1.26)	1.15 (1.00-1.32)	1.16 (1.02-1.32)	1.04 (0.85-1.26)
86-90	1.39 (1.18-1.65)	1.35 (1.00-1.84)	1.41 (1.10-1.82)	1.44 (1.01-2.05)
81-85	2.43 (1.90-3.10)	3.72 (2.48-5.60)	2.18 (1.49-3.19)	1.67 (0.97-2.90)
0-80	3.00 (2.41-3.74)	3.34 (2.33-4.81)	3.55 (2.55-4.95)	1.92 (1.11-3.32)
Respiratory rate (/min)				
Not registered*	1.00			
0-9	4.15 (2.96-5.82)	3.71 (2.37-5.81)	4.29 (2.38-7.74)	6.59 (2.02-21.50)
10-19	2.37 (2.01-2.79)	2.28 (1.82-2.87)	2.27 (1.73-2.97)	2.58 (1.58-4.19)
20-29	3.15 (2.66-3.73)	3.11 (2.45-3.96)	2.96 (2.25-3.90)	3.53 (2.16-5.76)
>30	5.16 (4.19-6.35)	5.59 (4.07-7.69)	5.61 (4.03-7.81)	5.20 (3.00-9.03)
Heart rate (beats/min)				
>125	2.53 (2.23-2.87)	2.69(2.23-3.24)	2.10 (1.71-2.59)	2.92 (2.11-4.02)
101-125	1.52 (1.36-1.68)	1.45(1.24-1.69)	1.55 (1.31-1.83)	1.37 (1.03-1.81)
76-100	1.10 (1.01-1.19)	1.00(0.88-1.13)	1.20 (1.05-1.37)	1.18 (0.96-1.46)
51-75*	1.00			
0-50	1.34 (1.08-1.65)	1.10(0.74-1.62)	1.25 (0.90-1.74)	2.11 (1.38-3.23)

Table 3. Adjusted Odds Ratios (AORs) for the association between vital signs and ICU/MCU admission by different age groups. (continued)

	Total cohort	18-65 years	66-80 years	>80 years
Temperature (°C)				
Not registered*	1.00			
≥40	0.50 (0.32-0.78)	0.53(0.29-0.97)	0.53 (0.25-1.14)	0.30 (0.07-1.28)
38-39	0.35 (0.30-0.41)	0.38(0.30-0.49)	0.37 (0.28-0.47)	0.28 (0.18-0.43)
35-37	0.64 (0.59-0.70)	0.63(0.56-0.71)	0.66 (0.58-0.76)	0.64 (0.51-0.79)
31-34	1.36 (1.00-1.84)	1.58(1.02-2.43)	1.14 (0.67-1.92)	1.41 (0.62-3.17)
0-30	3.95 (1.59-9.82)	5.90(1.36-25.66)	1.89 (0.36-9.95)	11.6 (1.64-81.7)

Abbreviations: ICU = intensive care unit, MCU= medium care unit
*= reference category in the regression analyses
The following potential confounders were entered in the model for through backward stepwise regression: age, gender, top ten chief complaints, hospital, treating specialty, $SpO_{2,}$ temperature, proxies of disease severity (triage level, fluid administration, Glasgow Coma Scale, blood gas analyses) and proxies of disease comorbidities/complexity (number of consultations, lab tests, performed radiological tests).

Younger patients had larger increases in predicted ICU/MCU admission compared to older patients with increasing DBP, RR, HR, and decreasing SpO_2 categories (supplementary file 5 and 6). In contrast, AORs changed to a similar extent among all age categories for changing vital signs.

Subgroup and sensitivity analyses
The AORs for mortality increased or decreased to the same extent in the subgroups with suspected infection and trauma patients, and in the sensitivity analyses, regardless of the patient's age category (supplementary file 3, 4 and 7).

DISCUSSION

The present study has two main findings. First, in-hospital mortality increased gradually with decreasing SBP and SpO_2, without the existence of a single threshold for these vital signs. For DBP, MAP and HR, we found quasi-U-shaped associations with in-hospital mortality. A single cut-off existed for MAP, RR, and temperature, however, using these single cut-offs would ignore further increase of risk with more extreme values of these vital signs. Secondly, absolute mortality increased substantially more in older compared to younger patients with changing vital sign categories.

The use of a single threshold for SBP in many acute care guidelines and risk stratification tools deserves scrutiny.[23-25, 47, 48] The linear association between SBP and mortality in an unselected ED population corresponds with previous studies in traumatic brain injury, sepsis, hypertension and unselected ED patients.[28, 49, 50] Our study suggests that a single cut-off is also absent for SpO_2 correspondent to earlier findings in a study by Ljunggren et al. in 96,512 unselected ED patients.[11] Interestingly, temperatures above 35.5°C, i.e. normal temperature and fever, were not significantly associated with mortality in our and Ljunggren's study. This finding is in contrast to what has been suggested in several risk stratification tools, like the national and modified EWS, in which points are assigned for high temperature.[51] It should be noted, however, that the original EWS was not intended to be a predictor of outcome, but rather to be a simple tool to assist inexperienced nursing or medical staff in recognizing patients at risk for deterioration and in securing immediate, more experienced help.[52]

In addition to a threshold for temperature, we also found a single cut-off for MAP and RR. RR was the least registered vital sign, but our findings confirm the importance of RR given the strong association with mortality and the higher odds for ICU/MCU admission if RR was registered, compared to if it

was not registered. The threshold we found for MAP <70mmHg was similar to the thresholds used in several guidelines.[39, 40] Interestingly, the odds for ICU admission only started to increase with MAP below 60mmHg. Probably, physicians are less triggered to admit patients to an ICU or MCU with a MAP between 60-80mmHg. This discrepancy between the associations of MAP and in-hospital mortality versus ICU admission warrants further research as it may reflect unrecognized hypotension, and patients with higher MAPs may need ICU admission.

Different from other studies, we have studied the effect of age on associations between vital signs and outcome. Older patients had larger increases in the absolute risk for mortality with changing vital sign categories compared to younger patients, while the relative risk for mortality had similar increases. Previously, the use of different single cut-off values for SBP in trauma patients has been suggested for advanced age.[53] Although prognosis has been found to deteriorate at higher SBP threshold, our data suggest not to use a single cut-off at all.

The findings of the present study have several implications for clinical practice. Although a single cut-off may be convenient and easy to use, our first finding suggests that the use of a cut-off for several vital signs in risk stratification tools and acute care guidelines are inappropriate.[21-23, 39-41] The use of a single cut-off suggests that above or below this specific threshold prognosis changes substantially. Using only a single threshold, like for SBP or RR in qSOFA, does not mean that the chosen threshold for SBP of 100mmHg discriminates better between low and high-risk patients than a threshold of 90, 110, 120, 130, 140 mmHg and so on, as has been demonstrated previously for SBP in traumatic brain injury.[28] Even if a single cut-off exists, i.e. if an association between outcome and a vital sign is only apparent from a certain value, dichotomous risk stratification would ignore further increase of the risk with more extreme values of the vital sign. This implicates that several risk stratification tools and acute care guidelines could be improved by using numerical scores based on AORs or predicted mortality rather than using dichotomous variables.[21-23, 39-41] Currently, these grading scores are already used in EWS, although the scores in the EWS are not supported by AORs derived from prediction models.[51]
APACHE II is a good example of a numerical score for ICU patients.[54] Similar risk stratification scores may be developed for the ED to improve recognition, initial resuscitation and disposition.

Our second main finding implicates that risks scores need to be age-adjusted. The effect of vital signs on mortality did change with advancing

age, because baseline risk for older age was higher. For example, if SBP decreased and the risk on in-hospital mortality increased fourfold, the effect on younger patients with a baseline mortality risk of i.e. 0.5% would be much smaller than the effect on older patients with a baseline mortality risk of i.e. 4.0%. More points may be assigned in a numerical score for older age for similar vital signs.
Interestingly, the subgroup analyses in sepsis and trauma patients suggests that the etiology of changes in vital signs does not affect the impact of vital signs on relevant outcomes. Future studies should incorporate our findings in development of risk stratification tools and acute care guidelines for ED patients using numerical scores based on absolute risks.

Our study has several strengths like a large sample size with unselected patients and the multicentre design. Several limitations merit emphasis. First, this observational study used the NEED, which lacks information about the medical history and comorbidities of patients. However, to overcome this, we used variables that are associated with comorbidities and complexity.[45] Second, information bias cannot be completely ruled out although data collection was largely automatized and only reliably registered variables are collected in the NEED. Finally, we missed outcomes of patients who were hospitalized elsewhere. However, the sensitivity analyses in which we considered these patients as deceased, yielded similar results.

CONCLUSION

The use of a single cut-off for each vital sign in acute care guidelines and risk stratification tools deserves scrutiny, as our results show that clinical outcomes increase gradually or in a U-shaped fashion with changing vital signs. A threshold for MAP, RR and temperature exists, however, using these single cut-offs would ignore further increase of risk with more extreme values of these vital signs. Age-adjusted numerical scores will improve risk stratification, as older patients have larger absolute increases in mortality with changing vital signs, even after adjusting for confounders.

SUPPLEMENTARY FILE 1

Data collected in the Netherlands Emergency department Evaluation Database (NEED) and their definitions are described in detail.

GENERAL DATA OF PARTICIPATING EMERGENCY DEPARTMENTS

In the NEED, several hospitals and the emergency department characteristics are registered once a year.

1. Which hospital. Coding: <Hospital Code>.

2. Type of hospital. Urban = 0. Urban (STZ = Foundation of top clinical hospitals) teaching hospital = 1. Academic Medical Centre = 2.

3. Number of patients in the region of the relevant hospital. This information can be obtained for each participating hospital by the NZa (was once used to calculate the hospital budgets).

4. Emergency physicians. a. Number of emergency physicians working at the emergency room in the hospital on the first of January of the year in question. b. Emergency physicians in fulltime equivalent. c. Number of residents in emergency medicine on the first of January of the year in question. d. Number of residents not in training on the first of January of that year.

5. Presence of emergency physicians. a. Non-24/7 presence by emergency physicians = 0. Minimum of 1 emergency physician 24 hours a day, 7 days a week = 1. b. How many emergency physicians, residents in emergency medicine and residents not in training are working average at the emergency department daily for patient care. Number per shift. The number of shifts for emergency physicians also must be defined.

6. Number of trauma surgeons employed by the hospital (so NO general surgeons). Trauma surgeons in FTE.

7. Number of internists in acute medicine employed by the hospital. (So NOT general internists).

8a. Trauma centre No = 0. Level 2 = 1. Level 1 = 2.

8b. Intensive care in hospital. Not present = 0. Present = 1.

9. Medium Care (or high care) present in hospital. Not present = 0. Present = 1.

10. Number of visits per year on the emergency department (so NOT the number of emergency patients per year). This only accounts for the number of patients at the emergency department. If a separate emergency cardiac care is present or GP clinic is integrated, these patient visits are counted separately.

11. Number of emergency department visits per year. N.

12. General Practitioner clinic integrated within the emergency department. No = 0. Yes = 1.

13. Emergency cardiac care present AT or NEXT to the emergency department. Not present = 0. Present = 1

14. Quality indicators by the Dutch Society of Emergency Physicians (DSEP (NVSHA). See also the NVSHA website. www.nvsha.nl.

Quality indicators for Acute Coronary Syndrome (ACS).

1.Are patients with (suspected) ACS treated at your emergency department? Indicate whether any patients are suspected for acute coronary syndrome at your emergency department. Answer: No = 0. Yes = 1.

2. For patients with (suspected) ACS, do you have a protocol about administration of aspirin?

Indicate whether a protocol is used at the emergency department for the administration of aspirin in cases of suspected acute coronary syndrome. Mention which protocol. Answer: No = 0. Yes = 1.

3. Time for administration of aspirin to patients with (suspected) ACS registered at the emergency room?

Specify whether the exact time of administration of aspirin given to patients with suspicion of acute coronary syndrome is registered at the emergency department. This can be a written or electronic registration. Answer: No = 0. Yes = 1.

Quality indicator Complication registration.

A complication is an unintended and undesired event or condition during or following medical treatment which is harmful to the health of the patient needing change in medical treatment, or if irreversible health damage exists.

1. Complication registrations

Is a complication registration used at the emergency department? Only one answer is possible. If this question is answered with "no", the other questions with indicator "Triage on Emergency Department" do not apply. You can continue to answer the questions with the next indicator. Answer: No = 0. Yes = 1.

2. Complication meeting

A complication meeting at least four times a year at the emergency room? Answer only yes or no. Answer: No = 0 Yes = 1.

3. Complication Registration

Here we would like you to briefly specify which CR you are using. If possible, you can attach a copy of the list of complications registered. Please note: This does not concern completed lists with numbers of complications.

Free text: ...

Quality indicator child abuse / domestic violence.

1.Presence of child abuse protocols

A written child abuse protocol used at the emergency department which meets the following minimum competencies? A. A multidisciplinary child abuse team is active in the hospital. B. Employees at the emergency department are trained in identifying child abuse. Answer: Yes, to both: 3 A yes, B no: 2 A no, B yes: 1 No = 0.

2. Screening instrument

Does the ED use a screening instrument to signal child abuse, such as the SPUTOVAMO form or a derivative? Answer: No = 0. Yes = 1.

Which screening instrument is used at the emergency room?

(multiple answers possible): SPUTOVAMO form. No = 0. Yes = 1 Top - toe examination. No = 0. Yes = 1. Otherwise, describe: <Free text ...>

3. Number of completed screening documents.

What is the number of completed screening documents at the emergency department, in relation to the total number of children up to and including 18 years old at the emergency department? **n** Number of completed screening documents at the ED in January 2011

N Total number of children up to and including 18 years at the emergency room in January 2011

n / N * 100% Screening document percentage

4. Protocol "parent notifications" Is a protocol used at the emergency department for the reporting of children as victims of domestic violence / suicidal attempt / auto mutilation / excessive alcohol and / or drug use? Answer: No = 0. Yes = 1. Notifications:

Quality indicator pain relief.

1.Pain scoring systems Is a pain scoring system used at the emergency department? Answer: No = 0. Yes = 1. If yes, which scoring system: <Free text>

2. Time registration of pain relief. Is the exact time recorded when pain relief is given in the emergency room? Answer: No = 0. Yes = 1.

3. Pain protocols

Is a pain protocol used at the emergency room? Answer: No = 0. Yes = 1. Comments <Free text>

Quality indicator Sepsis.

1.Sepsis protocol Is a sepsis protocol used at the emergency department? Answer: No=0. Yes=1. If yes, which protocol <free text>

2. Screening document Is a screening document used for sepsis in the emergency room? Answer: No=0. Yes=1. If yes, which document: <free text>

3. Time registration for administering of antibiotics to patients with sepsis Is time of administering of antibiotics to patients with sepsis registered at the emergency department? Answer No=0. Yes=1. Comments <free text>

Quality indicator Procedural sedation and analgesia (PSA)

PATIENT RECORDS.

The following basic principles are used: 1) Facts are recorded, e.g. times. Derived variables are generated afterwards, such as duration of stay. 2) Continuous variables are not categorized because this will lead to data loss. 3) Only create categories for coding, considering a remaining category "other" and the possibility to indicate "unknown" The following data is collected from all patients:

Demographic information. In general, the social service number (BSN) and date of birth are registered for each patient. In addition, a unique code is registered for each emergency room visit.

1.Age Age in years at the time of registration at the emergency registration desk.

2. Sex. Male or female. Code: Female = 0. Male = 1

3. Date and time Registration of ED presentation. DD-MM-YEAR-HOUR-MIN.

Transport to the hospital.

4. Type of arrival at the hospital. Arrival at the emergency room by ambulance or by own transport (i.e. walking, bicycle / moped or public transport). Encoding: Own transport = 0. Arrival with ambulance = 1. 9 = Unknown

5. Referral status. Self-referrer or referred by GP or another specialist. Encoding: Self-referrer = 0. Referred by physician = 1. Referred by specialist = 2. 9 = Unknown. For variable 4 and 5, more variables are possible. This still has to be discussed and agreed with the participating hospitals. Example: Anyone who enters by ambulance after a 112-call, would that be a self-referrer?

General information in the ED.

6. Triage category. Triage category according to Manchester triage system upon arrival of patient. Coding: Blue = 1, green = 2, yellow = 3, orange = 4, red = 5. If the Boston triage system is used, the corresponding categories 1 to 5 will be used. If a Dutch triage system is used, it will be entered.

7. Main specialism or supervisor. In case this is not known, the specialty will be recorded which admits the patient.

Coding: emergency physician = 0. Surgery = 1. Internal medicine (or super specialism of internal medicine for example gastroenterologist or vascular medicine) = 2. Cardiology = 3. Neurology = 4. Urology, ENT or ophthalmology = 5. Pediatrics = 6. Other = 7.

8. Presenting complaint according to MTS, NTS, etc triage system. So not the entry complaint according to the registration desk employee. Coding: Complaint 1 = 1, complaint 2 = 2, ... Complaint 52 = complaint 52.

9. Resuscitation in shock (trauma) room. Coding: Treatment in standard treatment room = 0. Shock room / trauma room = 1.

The definition of a shock room still needs to be determined by the participating emergency departments. They need to agree whether, for example, thrombolysis is part of this or not. The exact conditions of this room must be specified?

10. Initial vital signs at the time of triage or in the ED treatment room. So NOT mentioned on arrival at the treatment room. This variable only needs to be registered if it has been measured. If not measured, the space is left empty.

Oxygen saturation as measured with a pulse oximeter used in the hospital (percentages without oxygen). If measured when oxygen is given, note the number of liters / min of oxygen. Systolic and diastolic blood pressure (mmHg).Heart rate measured with pulse oximeter (beats / min.).Temperature measured with an ear thermometer (degrees Celsius). Glascow coma scale (EMV score). A separate variable is made of each vital parameter, with limit values and units. The Early Warning Score can then be calculated.

11a. Additional blood tests on the emergency room. If blood has not been obtained and sent for analyses, a 0 will be entered. If blood has been sent for analyses, a 1 will be entered.

11b. If blood has been obtained, the following values are entered. Carefully check whether the units at all EDs are the same! Blood gas obtained. 0 = No. 1 = Yes. If yes: Venous = 0. Arterial = 1. pH ...PO2 (KPa), PCO2 (KPa), Bicarbonate (mmol / L), B.E. ... *Biochemistry:* Sodium

(mmol / L), K (mmol / L), Creatinine (µmol / L) Urea (mmol / L), ASAT (U / L), ALAT (U / L), yGT (U / L), AF (U / L), LDH (U / L), CK (U / L), High sensitive Troponin (ng / L), CRP (mg / L), Pro-BNP (mg / L), Procalcitonin (mg / L), Lactate (mmol / L) *Hematology:* Hb (mmol / L), Leucocytes ($x10^9$ / L), Platelets ($x10^{12}$ / L), D-dimer (mg / L), INR *Toxicology:* Ethanol (mg / L), Paracetamol (mg / L) *Blood cultures collected at the emergency room and sent for cultures:*
0 = No. 1 = Yes. Blood cultures should be at least one set (i.e. an aerobic and an anaerobic). *Urine collected at the emergency room and send for cultures:* 0 = No. 1 = Yes. *Urinary sediment collected at the emergency room:* 0 = No. 1 = Yes. *Urine toxicological analysis performed at the emergency department:* 0 = No. 1 = Yes.

12a. Additional X-ray diagnostics at the emergency room.
If no X-ray diagnostics has been performed at the emergency room, a 0 is entered. If a type of X-ray diagnostics has been performed at the emergency room, a 1 is entered. Definition: 0 should also be entered, if the patient is sent for an X-ray by the GP, and afterwards sent to the Emergency Department.

12b. If X-ray diagnostics have been performed, the type of examination must be recorded. Conventional X-ray, extremity. 0 = No. 1 = Yes. Conventional X-ray, divers (including chest X-ray). 0 = No, 1 = Yes. Abdominal ultrasound (including FAST). 0 = No, 1 = Yes. Ultrasound for deep vein thrombosis (DVT). 0 = No, 1 = Yes. Ultrasound, divers. 0 = No, 1 = Yes. Head CT. 0 = No, 1 = Yes. CT-Cervical spine. 0 = No, 1 = Yes. CT pulmonary embolism. 0 = No, 1 = Yes. CT aorta. 0 = No, 1 = Yes. CT chest-abdomen (trauma). 0 = No, 1 = Yes. Patients can receive multiple examinations at the emergency room.
13. Number of consultations at the emergency room. Coding: No consultations = 0. A consultation = 1. Two consultations = 2. Three consultations = 3, etc.

Outcomes
14. Emergency department lengths of stay. Duration in minutes. Discharge time of the emergency room minus the registration time at the emergency room. (So NOT the announcement or triage time).
15. Discharge destination (disposition) Home, to general ward, medium care (MC) or Coronary Care Unit (CCU) or intensive care (ICU). Died in ED. Transfer to another hospital. Scheduled outpatient clinic after ED visit and related to ED visit.
16. Hospital length of stay. In days. Discharge date minus recording date. If a patient at the emergency room is discharged home, the hospital duration is 0 days. If people are registered before midnight and leave the ED after midnight, this should be accounted as 0 days!
17. Hospital mortality. Coding: Leaving the hospital alive = 0. Died in ED or brought in dead (for example during CPR) = 1. Died in hospital before discharge = 2.
18. Return to ED with medical problem / complaints. Coding: No return with medical problems within 7 days after registration time on ED = 0. Return to the ED within a week after discharge with a medical problem which may or may not be related to the previous ED visit = 1. Return to the ED within a week after discharge with a medical problem which is clearly related to the previous ED visit = 2.

Otherwise = 3

19. ICD-10 code and diagnosis after discharge from the emergency room and hospital.

20. Diagnosis Treatment Code (DBC or DOT).

SUPPLEMENTARY FILE 2

The following variables were calculated or modified:

- Triage levels were considered equal among hospitals. Non-urgent triage levels (blue and green) were merged into one triage level, because of the low number of ED patients and events in the triage level blue.
- The Manchester Triage System (MTS) was used in the tertiary care centre, and consisted of 51 possible chief complaints. The Dutch Triage Standard (NTS) consisted of 50 possible triage chief complaints and was used in the two urban hospitals. The MTS and NTS were merged into one chief complaints list. The MTS, NTS and the merged chief complaints list are shown below. If chief complaints did not match, a new chief complaint was made. In total the merged list contained 51 different chief complaints. The top ten chief complaints were used in the present study to adjust the primary associations of interest (feeling unwell; abdominal pain; dyspnea; chest pain; extremity problems; collapse; trauma; palpitations; wounds; headache, and a miscellaneous category containing all the other chief complaints.
- Treating specialty was categorised to minimalize the number of categories and use as potential confounder: surgical (i.e. surgery, orthopaedics, ophthalmology, urology, ear-nose-throat), medical (internal medicine, cardiology, neurology, gastroenterologist, psychiatry, and remaining specialties) and emergency medicine if no other specialty had been involved. Specialities were assigned before ED treatment during patient registration. Self-referred patients and critically ill patients are generally assigned to the Emergency Medicine. If a patient is referred by a general practitioner, he/she is often referred to a specific specialty.

Merged chief complaints (N=51)	MTS chief complaints (N=51)	NTS chief complaints (N=50)
Abdominal pain	Abdominal pain in adults Abdominal pain in children	Abdominal pain in adults Abdominal pain in children
Abscesses & local infections	Abscesses and local infections	
Allergy, bites & stings	Allergy Bites and stings	Allergic reaction and stings
Apparently drunk	Apparently drunk	
Assault	Assault	
Asthma	Asthma	
Back pain	Back pain	Back pain
Behaving strangely & suicidal	Behaving strangely	Behaving strangely or suicidal
Breast infection		Breast infection
Burns & scalds	Burns and scalds	Burns and scalds
Chest pain	Chest pain	Chest pain
Collapse	Collapsed adult	Collapse Dizziness
Constipation		Constipation
Coughing		Coughing
Crying baby	Crying baby	
Dental problems	Dental problems	Dental problems
Diabetes	Diabetes	Diabetes
Diarrhea & vomiting	Diarrhea and vomiting	Diarrhea Vomiting

Merged chief complaints (N=51)	MTS chief complaints (N=51)	NTS chief complaints (N=50)
Dyspnea	Shortness of breath in adults Shortness of breath in children	Shortness of breath
Ear problems	Ear problems	Ear problems
Exposure to chemicals	Exposure to chemicals	
Extremity problems	Limb problems	Arm problems General/limb trauma Leg problems
Eye problems	Eye problems	Eye problems
Facial problems	Facial problems	Facial trauma Nosebleed
Falls	Falls	
Feeling unwell	Unwell adult Unwell child	Fever in adults Fever in children Neurological failure Unwell adult Unwell child
Fits	Fits	Fits
Foreign body	Foreign body	Foreign body
Gastro-intestinal (GI) bleeding	Gastro-intestinal (GI) bleeding	
Genital problems	Testicular pain	Genital problems
Headache	Headache	Headache
Implantable Cardioverter Defibrillator (ICD)		Implantable Cardioverter Defibrillator (ICD)

Merged chief complaints (N=51)	MTS chief complaints (N=51)	NTS chief complaints (N=50)
Irritable child	Irritable child	
Limping child	Limping child	
Major incidents – primary	Major incidents – primary	
Mental illness	Mental illness	
Near-drowning		Near-drowning
Neck pain	Neck pain	Neck problems Neck trauma
Overdose & poisoning	Overdose and poisoning	Poisoning
Palpitations	Palpitations	Palpitations
Per vaginum (VP) bleeding	Per vaginum (VP) bleeding	Per vaginum (VP) bleeding
Pregnancy	Pregnancy	Childbirth
Rashes	Rashes	Rashes
Rectal problems		Rectal problems
Self-harm	Self-harm	
Sexually acquired infection	Sexual acquired infection	
Throat problems	Sore throat	Throat problems

Merged chief complaints (N=51)	MTS chief complaints (N=51)	NTS chief complaints (N=50)
Trauma	Head injury Torso injury Major trauma	Head trauma Thorax trauma Abdominal trauma Back trauma
Urinary problems	Urinary problems	Urinary problems
Worried parent	Worried parent	
Wounds	Wounds	Wounds

Abbreviations: MTS=Manchester Triage System, NTS= Dutch Triage Standard, N= number

SUPPLEMENTARY FILE 3

Adjusted Odds Ratios (AORs) for the association between In-hospital mortality and vital signs by different age groups in a subgroup analysis: suspected infection.

	Total cohort	18-65	66-80	>80
Systolic blood pressure (mmHg)				
>140*	1.00			
121-140	1.01 (0.79-1.28)	.97 (0.53-1.77)	1.08 (0.75-1.54)	0.96 (0.65-1.41)
101-120	1.56 (1.24-1.96)	1.74 (1.02-2.98)	1.63 (1.15-2.33)	1.41 (0.96-2.06)
81-100	1.85 (1.42-2.40)	1.39 (0.74-2.63)	2.49 (1.68-3.68)	1.64 (1.05-2.57)
0-80	2.58 (1.88-3.56)	3.55 (1.82-6.93)	2.62 (1.55-4.42)	2.08 (1.17-3.69)
Diastolic blood pressure (mmHg)				
>120*	1.00			
101-120	0.60 (0.36-1.03)	1.21 (0.40-3.60)	0.50 (0.23-1.10)	0.40 (0.15-1.09)
81-100	0.68 (0.43-1.08)	0.68 (0.25-1.91)	0.54 (0.28-1.04)	0.89 (0.39-2.00)
61-80	0.64 (0.41-1.01)	0.91 (0.33-2.49)	0.50 (0.26-0.96)	0.71 (0.32-1.59)
0-60	1.00 (0.64-1.58)	1.59 (0.57-4.44)	0.91 (0.47-1.76)	0.89 (0.40-2.00)
Mean Arterial blood pressure (mmHg)				
>120*	1.00			
101-120	0.86 (0.63-1.18)	1.02 (0.48-2.16)	0.79 (0.49-1.26)	0.85 (0.50-1.45)
81-100	0.89 (0.66-1.19)	1.12 (0.56-2.27)	0.71 (0.45-1.11)	0.98 (0.60-1.62)
61-80	1.22 (0.90-1.67)	1.33 (0.64-2.77)	1.35 (0.86-2.14)	1.07 (0.63-1.83)
0-60	2.22 (1.52-3.23)	3.23(1.38-7.58)	2.34 (1.31-4.17)	1.75 (0.92-3.33)

	Total cohort	18-65	66-80	>80
Peripheral oxygen saturation (%)				
96-100*	1.00			
91-95	1.09 (0.91-1.31)	1.15 (0.76-1.75)	1.08 (0.82-1.42)	1.01 (0.74-1.36)
86-90	1.37 (1.04-1.81)	1.55 (0.79-3.05)	1.43 (0.96-2.12)	1.18 (0.72-1.94)
81-85	2.01 (1.32-3.07)	1.10 (0.38-3.22)	2.40 (1.26-4.59)	2.09 (1.03-4.24)
0-80	3.24 (2.13-4.86)	2.80 (1.21-6.45)	3.78 (2.02-7.09)	2.58 (1.25-5.35)
Respiratory rate (/min)				
Not measured*	1.00			
0-9	#	#	#	#
10-19	1.08 (0.78-1.49)	1.19 (0.61-2.33)	0.95 (0.59-1.54)	1.11 (0.60-2.05)
20-29	1.66 (1.20-2.27)	2.20 (1.13-4.31)	1.48 (0.93-2.34)	1.52 (0.84-2.76)
>30	2.45 (1.70-3.51)	3.86 (1.73-8.60)	2.61 (1.54-4.43)	1.88 (0.98-3.62)
Heart rate (beats/min)				
>125	2.63 (1.88-3.68)	4.03 (1.88-8.66)	1.70 (0.99-2.93)	2.99 (1.70-5.26)
101-125	1.78 (1.34-2.37)	1.59 (0.76-3.33)	1.68 (1.09-2.60)	2.05 (1.29-3.28)
76-100	1.33 (1.01-1.74)	1.55 (0.76-3.14)	1.13 (0.75-1.72)	1.57 (1.02-2.43)
51-75*	1.00			
0-50	1.64 (0.89-2.99)	3.52 (0.83-15.0)	1.12 (0.40-3.17)	2.12 (0.84-5.31)

	Total cohort	18-65	66-80	>80
Temperature (°C)				
Not measured*	1.00			
≥40	0.41 (0.23-0.75)	0.35 (0.10-1.25)	0.35 (0.13-0.95)	0.53 (0.20-1.38)
38-39	0.52 (0.39-0.68)	0.59 (0.31-1.11)	0.54 (0.36-0.80)	0.41 (0.26-0.65)
35-37	1.14 (0.88-1.48)	1.50 (0.82-2.72)	1.01 (0.69-1.49)	1.09 (0.70-1.71)
31-34	2.59 (1.50-4.46)	5.97 (2.18-16.4)	1.28 (0.47-3.51)	2.41 (0.95-6.15)
0-30	0.70 (0.08-6.26)	#	#	#

*= reference category, # No outcome events for this category
Case definition: Patients with suspected infection were selected if blood cultures were obtained in the ED, and if patients were hospitalized.

SUPPLEMENTARY FILE 4

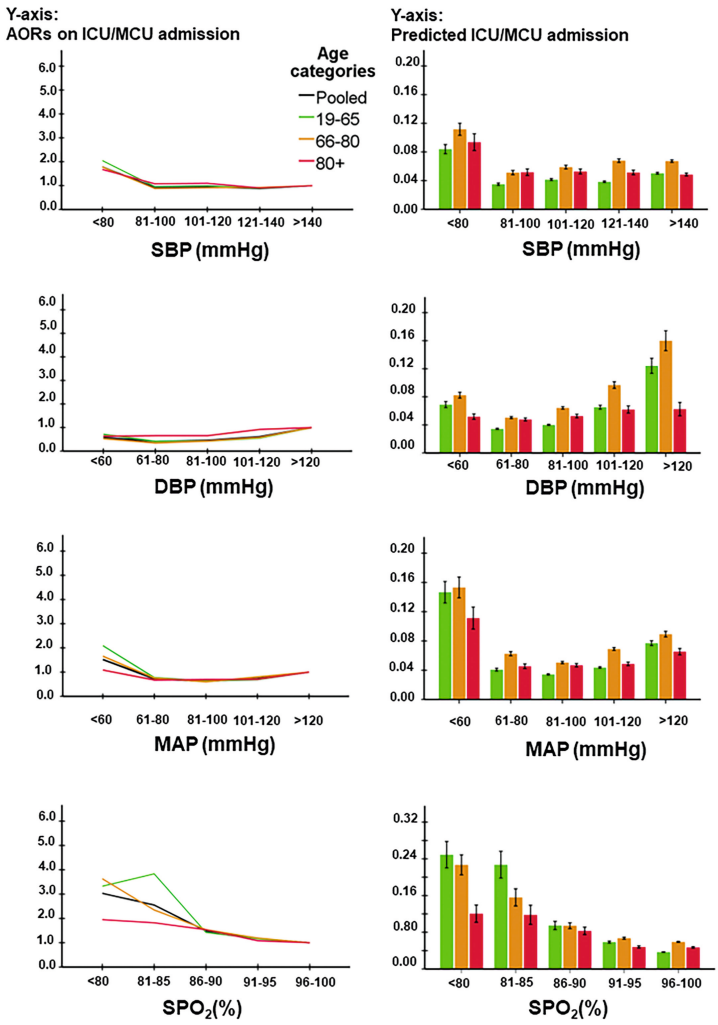
Adjusted Odds Ratios (AORs) for the association between In-hospital mortality and vital signs in a subgroup analysis: Trauma.

	Total cohort
Systolic blood pressure (mmHg)	
>140*	1.00
121-140	1.18 (0.73-1.94)
101-120	1.31 (0.75-2.30)
81-100	1.64 (0.79-3.41)
0-80	1.72 (0.67-4.40)
Diastolic blood pressure (mmHg)	
>120*	1.00
101-120	0.80 (0.29-2.23)
81-100	0.35 (0.13-0.91)
61-80	0.46 (0.18-1.20)
0-60	0.72 (0.26-2.01)
Mean Arterial blood pressure (mmHg)	
>120*	1.00
101-120	0.50 (0.30-0.85)
81-100	0.56 (0.33-0.95)
61-80	1.13 (0.63-2.04)
0-60	1.70 (0.64-4.50)
Peripheral oxygen saturation (%)	
96-100*	1.00
91-95	1.87 (1.24-2.82)
86-90	5.75 (3.14-10.52)
81-85	6.88 (2.61-18.12)
0-80	4.59 (1.55-13.57)
Respiratory rate (/min)	
Not measured*	1.00
0-9	#
10-19	1.95 (1.04-3.65)
20-29	2.28 (1.16-4.48)
>30	4.55 (1.35-15.31)

	Total cohort
Heart rate (beats/min)	
>125	4.22 (1.76-10.1)
101-125	2.49 (1.43-4.33)
76-100	1.03 (0.66-1.59)
51-75*	1.00
0-50	2.05 (0.74-5.68)
Temperature (°C)	
Not measured*	1.00
≥40	#
38-39	1.16 (0.42-3.19)
35-37	1.20 (0.78-1.85)
31-34	4.84 (1.13-20.75)
0-30	0.74 (0.04-15.27)

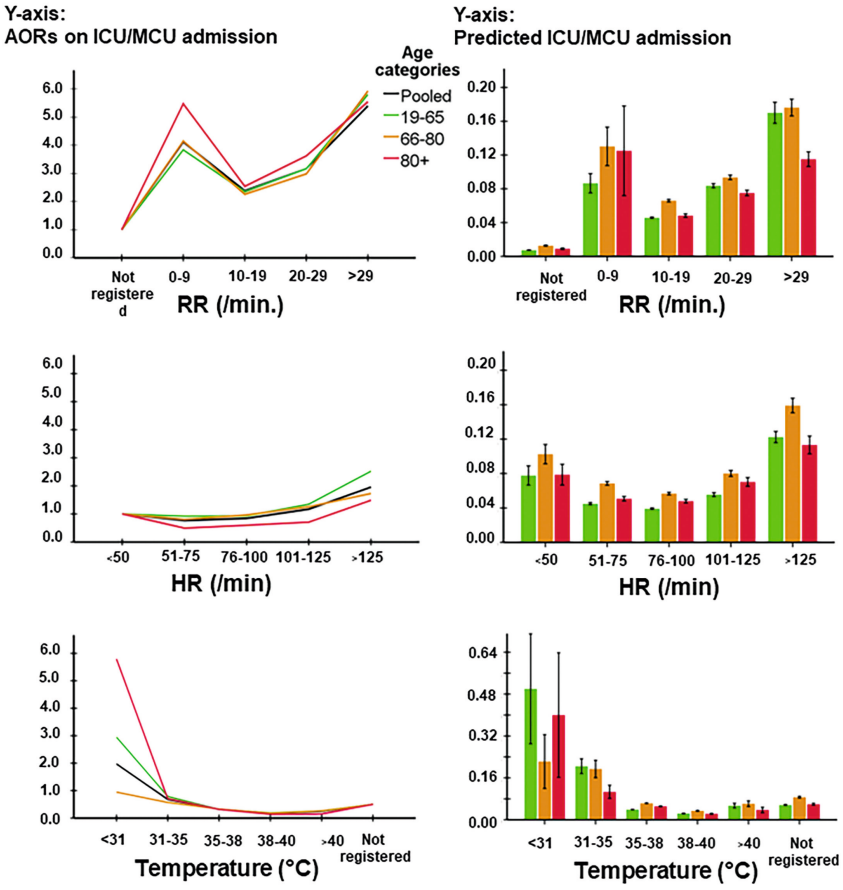
*= reference category
No outcome events for this category
Case definition: Trauma patients were selected by hospitalized patients, triaged orange or red and by the following triage presenting complaints: Severe trauma; facial complaints; fall; extremity complaints; large scale incidents; neck pain; thorax injury; head injury; wounds. Additionally, a radiological test had to be performed.

SUPPLEMENTARY FILE 5



Adjusted odds ratios (AORs) for systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP) and peripheral oxygen saturation (SpO_2) and intensive care unit (ICU) / medium care unit (MCU) admission (left side) and predicted ICU/MCU admission (right side) as a function of vital signs in different age categories. Note that the black line is the pooled data of all three age categories together. The following potential confounders were entered in the model for in-hospital mortality through backward stepwise regression: age, gender, triage level, top ten chief complaints, hospital, treating specialty, fluid administration, SpO_2, temperature, Glasgow Coma Scale, lab tests, number of consultations in the ED and performed radiological test.

SUPPLEMENTARY FILE 6



Adjusted odds ratios (AORs) for respiratory rate (RR), heart rate (HR) and temperature for intensive care unit (ICU) / medium care unit (MCU) admission (left side) and predicted ICU/MCU admission (right side) as a function of vital signs in different age categories. Note that the black line is the pooled data of all three age categories together. The following potential confounders were entered in the model for in-hospital mortality through backward stepwise regression: age, gender, triage level, top ten chief complaints, hospital, treating specialty, fluid administration, SPO_2, temperature, GCS, lab tests, number of consultations in the ED and performed radiological test.

SUPPLEMENTARY FILE 7

Sensitivity analysis in which the associations between vital signs and in-hospital mortality were investigated. Patients transferred to other hospitals from the ED were considered as deceased.

	Total cohort
Systolic blood pressure (mmHg)	
>140*	1.00
121-140	1.14 (1.03-1.26)
101-120	1.63 (1.47-1.80)
81-100	1.72 (1.52-1.95)
0-80	1.98 (1.69-2.32)
Diastolic blood pressure (mmHg)	
>120*	1.00
101-120	0.73 (0.60-0.90)
81-100	0.67 (0.56-0.81)
61-80	0.71 (0.59-0.85)
0-60	0.97 (0.80-1.17)
Mean Arterial blood pressure (mmHg)	
>120*	1.00
101-120	0.86 (0.77-0.97)
81-100	0.98 (0.87-1.10)
61-80	1.22 (1.08-1.39)
0-60	1.76 (1.46-2.12)
Peripheral oxygen saturation (%)	
96-100*	1.00
91-95	1.31 (1.21-1.43)
86-90	1.92 (1.66-2.23)
81-85	2.58 (2.06-3.23)
0-80	3.97 (3.26-4.84)
Respiratory rate (/min)	
Not measured*	1.00
0-9	0.84 (0.50-1.41)
10-19	1.15 (1.03-1.29)
20-29	1.60 (1.42-1.81)
>30	2.56 (2.16-3.03)

	Total cohort
Heart rate (beats/min)	
>125	1.89 (1.61-2.22)
101-125	1.59 (1.43-1.78)
76-100	1.12 (1.03-1.22)
51-75*	1.00
0-50	1.27 (1.01-1.60)
Temperature (°C)	
Not measured*	1.00
≥40	0.38 (0.23-0.60)
38-39	0.60 (0.52-0.68)
35-37	0.86 (0.79-0.94)
31-34	2.05 (1.53-2.75)
0-30	1.00 (0.36-2.81)

*= reference category
No events for this category


EMERGENCY

3

AGE-ADJUSTED INTERPRETATION OF BIOMARKERS OF RENAL FUNCTION AND HOMEOSTASIS, INFLAMMATION, AND CIRCULATION IN EMERGENCY DEPARTMENT PATIENTS.

Bart GJ Candel
Jamèl Khoudja
Menno I Gaakeer
Ewoud ter Avest
Ozcan Sir, Heleen Lameijer
Roger APA Hessels
Resi Reijnen
Erik W van Zwet
Evert de Jonge
Bas de Groot

Published in Nature Scientific Reports – 2022, Volume 12, issue 1, p1-9.

ABSTRACT

Background Appropriate interpretation of blood tests is important for risk stratification and guidelines used in the Emergency Department (ED) (such as SIRS or CURB-65). The impact of abnormal blood test values on mortality may change with increasing age due to (patho)-physiologic changes. The aim of this study was therefore to assess the effect of age on the case-mix adjusted association between biomarkers of renal function and homeostasis, inflammation and circulation and in-hospital mortality.

Methods Observational multi-center cohort study using the Netherlands Emergency department Evaluation Database (NEED), including all consecutive ED patients ≥18years of three hospitals. A generalized additive logistic regression model was used to visualize the association between in-hospital mortality, age and five blood tests (creatinine, sodium, leukocytes, C-reactive Protein, and hemoglobin). Multivariable logistic regression analyses were used to assess the association between the number of abnormal blood test values and mortality per age category (18-50;51-65;66-80;>80 years).

Results Of the 94,974 included patients, 2550 (2.7%) patients died in-hospital. Mortality increased gradually for C-reactive Protein (CRP), and had a U-shaped association for creatinine, sodium, leukocytes, and hemoglobin. Age significantly affected the associations of all studied blood tests except for leukocytes. In addition, with increasing age categories, case-mix adjusted mortality increased with the number of abnormal blood tests.

Conclusions The association between blood tests and (adjusted) mortality depends on age. Mortality increases gradually or in a U-shaped manner with increasing blood test values. Age-adjusted numerical scores may improve risk stratification. Our results have implications for interpretation of blood tests and their use in risk stratification tools and acute care guidelines.

BACKGROUND

In Emergency Department (ED) patients, correct risk stratification is important to early recognize clinical deterioration and to appropriately assign patients to a definitive level of care.[55] Many risk stratification tools, such as SIRS, CURB-65, PIRO, GRACE, Ranson criteria, assign points for abnormal blood test values.[21, 22, 39, 56-59] A previous study showed, however, that risk stratification tools are ineffective at predicting mortality in older ED patients.[27] This may be caused by a different impact of abnormal blood test values on mortality with increasing age due to (patho)-physiological changes.[60, 61] For example, older patients normally have higher creatinine levels than younger patients.[62] Thus, in younger patients an increase in plasma creatinine reflects a larger decrease in renal function compared to older patients and may therefore also carry a larger risk.

In addition, due to less physiological reserve, or higher frailty, with increasing age, an increasing number of abnormal blood test values may carry a larger risk in older patients.

As a result, using age-adjusted risk scores for blood tests may have the potential to improve risk stratification and consequently facilitate early recognition of clinical deterioration. This early recognition may potentially improve outcome of patients with medical conditions who need treatments that are time sensitive.

The aim of the present study is twofold: First, to assess the effect of age on the associations between biomarkers of renal function and homeostasis, inflammation and circulation and in-hospital mortality. Secondly, we investigate the risk of an increasing number of abnormal blood tests per age category.

METHODS

Study design and setting

This observational multi-center study was conducted in three EDs in the Netherlands, with each approximately 20,000-30,000 ED visits per year. Consecutive ED visits were registered of one tertiary care center (1 January 2017 – 8 June 2019), and two urban hospitals (1 January 2019 – 12 January 2020 and 1 January 2017 – 31 December 2019). This study was registered in the Netherlands Trial Register (ID NL8422) and approved by the medical ethics committee of the Leiden University Medical Center, the Netherlands.

All methods were performed in accordance to the principles outlined in the Declaration of Helsinki.

Selection of participants
All consecutive ED visits of ≥18 years were included in this study, if at least one of the blood tests described below was performed.

Data collection
Data were collected from the Netherlands Emergency department Evaluation Database (NEED), the national quality registry for EDs in the Netherlands (www.stichting.need.nl). Correspondent to other quality registrations, and in accordance with the "General Data Protection Act", an informed opt-out procedure was used in the participating hospitals for inclusion in the NEED. Detailed information about the NEED and collected data are available in an earlier publication.[63] We selected biomarkers which represent organ (dys-) function and inflammation because these biomarkers are known to have prognostic or diagnostic value in the ED, are commonly used in risk scores developed for the ED and are routinely assessed in the ED. For example, creatinine reflects renal function, which is the first organ to fail in early disease stages such as sepsis.[21, 22, 39, 56-59] The following routinely measured biomarkers were investigated: creatinine, urea, and sodium (reflecting renal function and homeostasis), C-reactive protein (CRP) and leukocytes (reflecting inflammation), lactate and hemoglobin (reflecting circulation). In Supplementary file 1, details of the used laboratory assays and equipment are described.

Outcome measures
In-hospital mortality (including death in the ED) was the primary outcome measure.

Sample size estimation
For objective 1, we aimed to adjust for 6 potential confounders in the multivariable logistic regression analysis. For objective 2, we aimed to adjust for 39 potential confounders as described in the main statistical analyses section. Approximately five to ten events per variable are needed to prevent overfitting in association studies.[44] The NEED contained 148,828 ED visits of patients ≥18 years of age. We estimated that in ~60% of the ED visits blood tests were performed resulting in ~90,000 ED visits which could be used for the analyses. Estimated in-hospital mortality would be ~3% of the overall population. Included patients were stratified by age, yielding ~90,000 / 4 age categories = 22,500 patients per age category. On average we would

have 0.03 x 22,500 = 675 events per group, sufficient to adjust for the 39 potential confounders.

Descriptive analysis
Patient characteristics were summarized per age-category (18-50, 51-65, 66-80, and >80 years),[62] as mean (SD) for normally distributed data, and median (IQR) for skewed data. Per age category, the percentage of blood test values outside of the reference range were reported.

Main statistical analyses
For objective 1, our goal was to study how creatinine, sodium, leukocytes, CRP, and hemoglobin are associated with mortality in emergency care, and how these associations change with age. We used the R package mgcv,[64] to fit a generalized additive logistic regression model (GAM) to the binary outcome in-hospital mortality. In this model, the log odds of the outcome depends in an arbitrary way on all six predictors (five blood parameters and age) and all their interactions. Clearly, such a complex model is overparameterized which could lead to poor, unstable performance if left unaddressed. However, it can be handled as part of the fitting procedure by using a quadratically penalized likelihood type approach. Effectively, this method enforces a smooth dependence of the log odds mortality on the six predictors. To visualize the association between mortality and the six predictors, we generated five graphs which present different views on the same model. In each of the graphs, we varied one of the predictors together with age, while the other four predictors were left constant at "typical" values. These were chosen as follows: creatinine= 80µmol/L, leukocytes= 8.0 x10^9/L, CRP= 10mg/L, Sodium= 140mmol/L, and hemoglobin= 9mmol/L.

This analysis was intended to be descriptive and exploratory, and therefore we did not attempt formal statistical inference in terms of p-values and confidence intervals.

Due to the complexity of the model, we could not add other biomarkers or potential confounders. To assess whether adjustment for other potential confounders affected our results, a second analysis was performed, in which we also studied the biomarkers urea and lactate. Multivariable logistic regression analyses were used with separate models for each biomarker for each age category. We simplified the model by categorizing blood tests, based on expected distribution and reference intervals (see Supplementary file 1). The following potential confounders were entered in the models through backward stepwise elimination: age, gender, triage level (non-urgent, urgent, very urgent, most urgent), top ten presenting complaints (supplementary file

2), systolic blood pressure, heart rate, peripheral oxygen saturation, hospital, high dependency care unit admission (yes or no), number of consultations in the ED (0, 1, 2 or >2) and performed radiological tests (0 or ≥1). The models for creatinine, urea, CRP, and leukocytes were adjusted for hemoglobin and sodium. The model for sodium was adjusted for hemoglobin and urea, and the model for hemoglobin was adjusted for sodium and urea. If variation inflation factors (VIF) were below three, multicollinearity was assumed not to be a problem. Triage level, high dependency care unit admission and vital signs were used to adjust for disease severity.[27] Vital signs were categorized in five or six categories to overcome non-linear associations, including a category 'not measured' to prevent missing data. In the NEED comorbidities were not registered. The number of consultations in the ED and the number of radiological tests were used as measure of comorbidities / complexity as described previously.[45] If variables were missing, the patient was excluded from the analyses. However, the used potential confounders had almost no missing data in the NEED. After the analyses for each biomarker was performed in all age categories, the analyses were repeated for the pooled data. An interaction term of age*blood test was added to study whether age affected the association between the studied biomarker and mortality. We considered age as an effect modifier if the p-value of the interaction term was <0.05. Adjusted odds ratio's (AORs) with 95% Confidence Intervals (95%-CI) and predicted probabilities (mean case-mix adjusted mortality) were reported to compare relative risk increases with absolute mortality with changing blood test values.

For objective 2, the number of abnormal blood test values, outside of their reference intervals, was calculated for each age category. The used reference intervals are mentioned in supplementary file 1. We studied the association between the number of abnormal blood tests and mortality with multivariable logistic regression using similar potential confounders as described above.

Data were analyzed using SPSS (Version 25.0) and R version 4.0 (packages foreign, dplyr, mgcv).

RESULTS

Patient inclusion and characteristics

Out of the 147,728 ED visits of patients ≥18 years, 94,974 patients were included in whom blood tests were performed (flow diagram in supplementary file 3). Mean age of included patients was 60.7 (19.0) years. In-hospital mortality per age category was: 153 (0.6%) patients in age category 18-50 years, 421 (1.8%) patients 51-65 years, 1086 (3.6%) 66-80 years, and 890 (6.3%) patients >80 years. Biomarkers were more often abnormal in older patients. For example, sodium was in 9.2% abnormal in age-category 18-50 years, and in 25.5% in age-category >80 years (difference 16.3%; 15.5-17.1%). See table 1 for all patient characteristics and supplementary file 4 for patient disposition and outcomes.

Table 1. Patient characteristics in the total cohort and per age category.

	Total cohort N=94,974 (100.0%)	18-50 years N=26,697 (28.1%)	51-65 years N=23,840 (25.1%)	66-80 years N=30,257 (31.9%)	>80 years N=14,180 (14.9%)
Demographics N, (%)					
Age, mean (SD)	60.7 (19.0)	35.5 (9.7)	58.5 (4.3)	72.9 (4.2)	85.7 (3.9)
Sex, female	47,082 (49.6)	14,979 (56.1)	11,013 (46.2)	13,297 (43.9)	7793 (55.0)
Hospital setting N, (%)					
Tertiary care centre	28,665 (30.2)	9396 (35.2)	7762 (32.6)	8448 (27.9)	3059 (21.6)
Top-10 presenting complaints N, (%)					
Collapse	4001 (4.2)	800 (3.0)	930 (3.9)	1510 (5.0)	761 (5.4)
Extremity problems	5803 (6.1)	1151 (4.3)	1278 (5.4)	1876 (6.2)	1498 (10.6)
Headache	2000 (2.1)	817 (3.1)	511 (2.1)	475 (1.6)	197 (1.4)
Palpitations	3150 (3.3)	554 (2.1)	960 (4.0)	1301 (4.3)	335 (2.4)
Chest pain	11,372 (12.0)	2864 (10.7)	3611 (15.1)	3645 (12.0)	1252 (8.8)
Wounds	1390 (1.5)	430 (1.6)	365 (1.5)	432 (1.4)	163 (1.1)
Feeling unwell	21,268 (22.4)	4095 (15.3)	5417 (22.7)	7801 (25.8)	3955 (27.9)
Abdominal pain	14,745 (15.5)	6875 (25.8)	3643 (15.3)	3197 (10.6)	1030 (7.3)
Dyspnea	12,126 (12.8)	2055 (7.7)	2882 (12.1)	4831 (16.0)	2358 (16.6)
Trauma	2449 (2.6)	859 (3.2)	507 (2.1)	639 (2.1)	444 (3.1)
Miscellaneous	16,670 (17.6)	6197 (23.2)	3736 (15.7)	4550 (15.0)	2187 (15.4)

Table 1. Patient characteristics in the total cohort and per age category. (continued)

	Total cohort N=94,974 (100.0%)	18-50 years N=26,697 (28.1%)	51-65 years N=23,840 (25.1%)	66-80 years N=30,257 (31.9%)	>80 years N=14,180 (14.9%)
Proxies for disease Severity N, (%)					
Triage level					
Blue/Green	18,157 (19.1)	5626 (21.4)	4286 (18.3)	5438 (18.2)	2807 (20.1)
Yellow	41,499 (43.7)	12,236 (46.6)	10.354 (44.1)	12,909 (43.3)	6000 (42.9)
Orange	28,436 (29.9)	7291 (27.7)	7367 (31.4)	9504 (31.9)	4274 (30.6)
Red	5449 (5.7)	1128 (4.3)	1472 (6.3)	1971 (6.6)	878 (6.3)
Biomarkers of renal function and homeostasis, Median (IQR), [N]b, {% outside reference range}					
Creatinine, µmol/L	77 (63-97)	69 (59-82)	75 (62-91)	83 (67-106)	91 (72-122)
Urea, mmol/L	5.8 (4.3-7.9)	4.3 (3.4-5.4)	5.5 (4.4-7.0)	6.7 (5.2-9.0)	8.1 (6.2-11.3)
Sodium, mmol/L	140 (137-142)	140 (138-142)	140 (137-142)	139 (136-141)	139 (136-141)
Biomarkers of inflammation, Median (IQR), [N], {% outside reference range}					
CRP, mg/L	10.4 (4.2-47)	7.3 (3.8-32)	10.3 (4.0-49)	12.0 (5.2-50)	12.0 (5.2-50)
Leucocytes, x10^9/L	9.1 (7.0-12.0)	9.3 (7.2-12.1)	8.9 {6.9-11.9)	9.0 (6.9-12.)	9.2 (7.1-12.2)
Biomarkers of Circulation Median (IQR), [N], {% outside reference range}					
Hemoglobin, mmol/L	8.4(7.6-9.2)	8.7 (7.9-9.3)	8.6 (7.8-9.3)	8.3 (7.4-9.1)	7.9 (7.1-8.7)
Lactate, mmol/L	1.6 (1.1-2.3)	1.5 (1.0-2.2)	1.5 (1.1-2.3)	1.6 (1.1-2.4)	1.6 (1.1-2.4)
Proxies for comorbidity and complexity N, (%)					
Radiological test*	58,780 (61.9)	13,713 (51.4)	14,503 (60.8)	19,922 (65.8)	10,642 (75.0)

Table 1. Patient characteristics in the total cohort and per age category. (continued)

	Total cohort N=94,974 (100.0%)	18-50 years N=26,697 (28.1%)	51-65 years N=23,840 (25.1%)	66-80 years N=30,257 (31.9%)	>80 years N=14,180 (14.9%)
Number of consultations					
None	32.791 (34.5)	10,600 (39.8)	8355 (35.0)	9649 (31.9)	4187 (29.5)
1	54.814 (57.7)	14,233 (53.4)	13,723 (57.6)	18,215 (60.2)	8643 (61.0)
2	6293 (6.6)	1596 (6.0)	1491 (6.3)	2046 (6.8)	1160 (8.2)
>2	929 (1.0)	223 (0.8)	221 (0.9)	306 (1.0)	179 (1.3)

Abbreviations: N= number, SD = standard deviation, GP = General Practitioner, IQR = interquartile range, CRP = C-Reactive Protein, µmol/L = micromole per Litre, mmol/L= millimole per litre, mg/L = milligram per litre.
Blood tests used in risk stratification and acute care guidelines were selected, reflecting biomarkers of renal function and homeostasis, inflammation, and circulation.
Number of patients in the total cohort was for creatinine (N=89,784), urea (N=88,816), sodium (N=91,617) and lactate (N=13,717), leukocytes (N=91,136), CRP (N=78,085) and haemoglobin (N=92,304).
* If one or more of the following radiological tests were performed: Ultrasound, radiography, and computer-tomography.

Main results

For objective 1, the associations between biomarkers and mortality are represented in figure 1 and figure 2. Creatinine, sodium, leukocytes, and hemoglobin had U-shaped associations with mortality, while mortality increased gradually with increasing CRP. In these figures it is also shown that the absolute risk for mortality increased more in older compared to younger patients with abnormal plasma values.

In the additional analyses, in which we also studied urea and lactate, relative risks (AORs) were presented (see table 2). Both biomarkers of renal function (creatinine and urea) had a more marked increase in AORs in patients 18-50 years with higher plasma concentrations. AORs started to increase below 136-145mmol/L in 18-50 years and below 126-130 mmol/L in >80 years. Mortality increased gradually for CRP, with more marked increase in patients 18-50 years for plasma levels of 51-100mg/L. Hemoglobin had a U-shaped association with mortality especially in older patients. A low hemoglobin was associated with higher mortality in younger compared to older patients. For lactate, mortality increased gradually in all age categories.

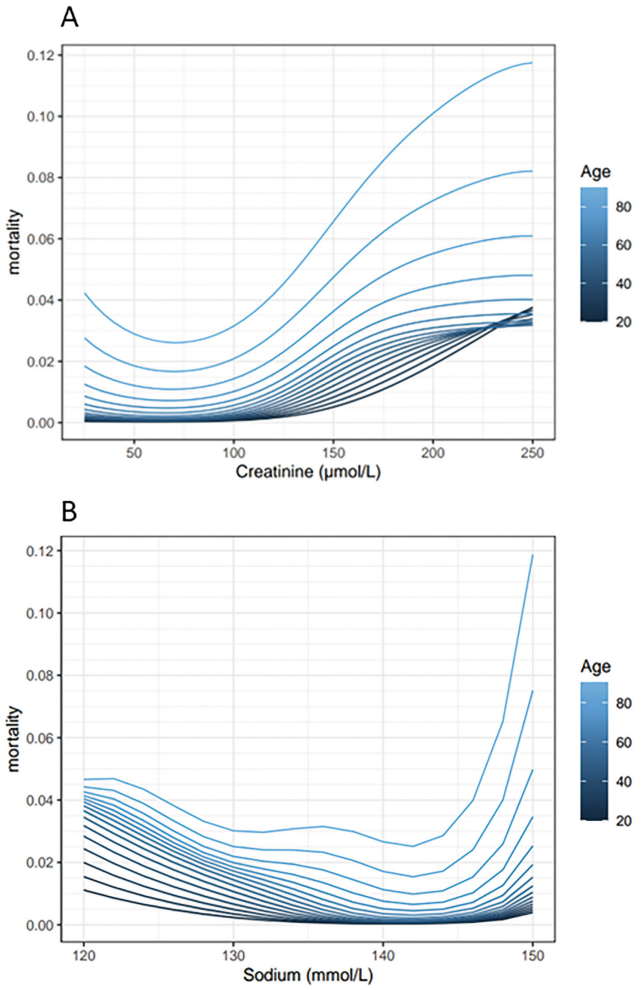


Figure 1. The associations between creatinine (panel A), sodium (panel B) and mortality and age are shown using a generalized additive logistic regression model. This model included five biomarkers and age. The four biomarkers that are not shown in the panel were left constant at 'normal' values. These were chosen as follows: creatinine= 80µmol/L, leukocytes=80x10^9/L, CRP=10mg/L, Sodium=140mmol/L, and hemoglobin=9mmol/L. Mortality is shown as predicted probability (between 0 and 1).

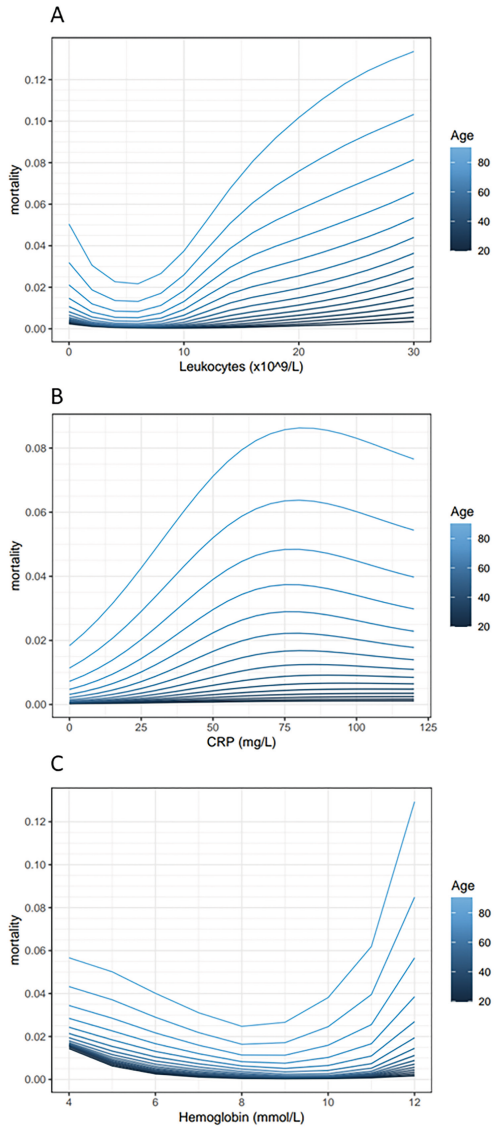


Figure 2 The associations between leukocytes (panel A), C-reactive Protein (panel B), hemoglobin (panel C) and mortality and age are shown using a generalized additive logistic regression model. This model included five biomarkers and age. The four biomarkers that are not shown in the panel were left constant at 'normal' values. These were chosen as follows: creatinine= 80µmol/L, leukocytes=8.0x10^9/L, CRP=10mg/L, Sodium=140mmol/L, and hemoglobin=9mmol/L. Mortality is shown as predicted probability (between 0 and 1).

Table 2 Adjusted odds ratios with 95% confidence intervals for the association of categorized urea, creatinine, sodium, leukocytes, C-Reactive Protein, haemoglobin and lactate and in-hospital mortality in age-categories.

	Total cohort AOR (95% CI)	18-50 years AOR (95% CI)	51-65 years AOR (95% CI)	66-80 years AOR (95% CI)	>80 years AOR (95% CI)	P-value
Biomarkers of renal function and homeostasis						
Urea, mmol/L						**0.03**
0-5.0*	1.0	1.0	1.0	1.0	1.0	
5.1-10.0	1.7 (1.5-2.0)	2.5 (1.5-4.0)	1.7 (1.3-2.3)	1.5 (1.2-1.9)	1.8 (1.2-2.6)	0.18
10.1-15.0	3.1 (2.6-3.7)	5.7 (2.6-12.2)	3.3 (2.2-5.0)	3.0 (2.3-4.0)	2.9 (1.9-4.3)	<0.01
>15.0	5.3 (4.4-6.3)	15.5 (8.3-29.1)	3.4 (2.3-5.0)	5.0 (3.8-6.5)	5.9 (4.0-8.7)	0.10
Creatinine, µmol/L						**<0.01**
0-50	1.5 (1.2-1.8)	1.5 (0.7-2.9)	1.5 (1.0-2.1)	1.4 (1.1-1.9)	1.5 (1.0-2.1)	0.35
51-100*	1.0	1.0	1.0	1.0	1.0	
101-150	1.5 (1.3-1.6)	2.9 (1.7-4.8)	1.9 (1.4-2.6)	1.4 (1.2-1.6)	1.3 (1.1-1.6)	**<0.01**
>150	2.6 (2.3-2.9)	5.2 (3.1-8.9)	2.3 (1.7-3.2)	2.4 (2.0-2.8)	2.7 (2.2-3.3)	**0.02**
Sodium, mmol/L						<0.01
>145	2.2 (1.8-2.8)	2.0 (0.9-4.4)	2.1 (1.2-3.7)	1.9 (1.3-2.8)	2.6 (1.8-3.8)	0.96
136-145*	1.00	1.0	1.0	1.0	1.0	
131-135	1.4 (1.3-1.6)	2.7 (1.6-4.5)	1.9 (1.5-2.6)	1.5 (1.2-1.8)	1.0 (0.8-1.2)	**<0.01**
125-130	1.7 (1.4-2.1)	2.5 (1.1-6.0)	2.1 (1.4-3.3)	2.0 (1.5-2.6)	1.2 (0.8-1.7)	**0.02**
<125	2.8 (2.2-3.7)	6.7 (2.6-17.2)	4.4 (2.5-7.6)	2.6 (1.7-4.0)	2.0 (1.2-3.3)	**0.01**

Table 2 Adjusted odds ratios with 95% confidence intervals for the association of categorized urea, creatinine, sodium, leukocytes, C-Reactive Protein, haemoglobin and lactate and in-hospital mortality in age-categories. (continued)

	Total cohort AOR (95% CI)	18-50 years AOR (95% CI)	51-65 years AOR (95% CI)	66-80 years AOR (95% CI)	>80 years AOR (95% CI)	P-value
Biomarkers of inflammation						
Leucocytes, x10^9/L						0.47
0-4.0	2.4 (2.0-3.0)	3.8 (2.7-8.6)	1.7 (1.1-2.8)	2.3 (1.7-3.1)	2.8 (1.8-4.4)	
4.1-8.0*	1.0	1.0	1.0	1.0	1.0	
8.1-12.0	1.3 (1.2-1.5)	1.3 (0.8-2.4)	1.2 (0.9-1.7)	1.3 (1.1-1.6)	1.4 (1.1-1.8)	
12.1-16.0	1.7 (1.5-2.0)	2.2 (1.2-4.1)	1.7 (1.2-2.4)	1.7 (1.4-2.1)	1.8 (1.4-2.2)	
16.1-20.0	2.2 (1.8-2.6)	2.1 (1.0-4.4)	2.6 (1.7-3.8)	2.0 (1.5-2.6)	2.2 (1.6-2.9)	
>20.0	2.7 (2.3-3.2)	4.9 (2.5-9.8)	2.1 (1.4-3.2)	2.5 (2.0-3.2)	3.0 (2.2-4.1)	
CRP, mg/L						**<0.01**
0-50*	1.0	1.0	1.0	1.0	1.0	
51-100	1.5 (1.3-1.7)	2.4 (1.3-4.3)	2.0 (1.4-2.8)	1.7 (1.4-2.1)	0.9 (0.7-1.2)	**<0.01**
101-150	1.9 (1.6-2.3)	1.8 (0.8-4.0)	1.8 (1.2-2.7)	2.2 (1.7-2.8)	1.7 (1.3-2.3)	0.77
>150	2.4 (2.1-2.7)	2.4 (1.3-4.4)	2.9 (2.2-4.0)	2.4 (2.0-2.9)	2.1 (1.7-2.7)	0.21
Biomarkers of circulation						
Haemoglobin, mmol/L						**<0.01**
>9.0*	1.0	1.0	1.0	1.0	1.0	
7.1-9.0	1.1 (1.0-1.2)	1.0 (0.6-1.5)	1.4 (1.0-1.9)	1.2 (1.0-1.4)	0.8 (0.7-1.0)	0.18
6.1-7.0	1.5 (1.3-1.7)	2.2 (1.1-4.1)	2.1 (1.4-3.0)	1.4 (1.1-1.8)	1.2 (0.9-1.5)	**<0.01**

Table 2 Adjusted odds ratios with 95% confidence intervals for the association of categorized urea, creatinine, sodium, leukocytes, C-Reactive Protein, haemoglobin and lactate and in-hospital mortality in age-categories. (continued)

	Total cohort AOR (95% CI)	18-50 years AOR (95% CI)	51-65 years AOR (95% CI)	66-80 years AOR (95% CI)	>80 years AOR (95% CI)	P-value
<6.0	1.7 (1.4-2.0)	2.8 (1.5-5.4)	2.2 (1.5-3.3)	1.8 (1.4-2.4)	1.1 (0.8-1.5)	**<0.01**
Lactate, mmol/L						
0-2.0*	1.0	1.0	1.0	1.0	1.0	**<0.01**
2.1-4.0	1.7 (1.4-2.1)	3.4 (1.2-9.6)	2.3 (1.3-4.1)	1.5 (1.1-2.2)	1.7 (1.2-2.5)	0.31
4.1-6.0	3.3 (2.4-4.5)	8.0 (2.0-32.3)	4.2 (1.8-9.7)	2.4 (1.4-3.9)	4.1 (2.4-7.1)	0.44
>6.0	10.2 (7.6-13.4)	30.7 (11.0-85.6)	14.7 (7.4-28.9)	8.5 (5.4-13.5)	8.7 (4.7-15.9)	**<0.01**

Abbreviations: AOR = Adjusted Odds Ratio, 95-CI = 95% confidence interval, CRP = C-reactive Protein.
The following potential confounders were entered in the model through backward stepwise regression: age, gender, triage category (green/blue, yellow, orange, red), top ten presenting complaints, hospital, systolic blood pressure, heart rate, peripheral oxygen saturation, high dependency care unit admission, number of consultations in the ED (0, 1, 2 or >2), performed radiological tests and blood tests.
Number of patients in the total cohort was for creatinine (N=89,784), urea (N=88,816), sodium (N=91,617) and lactate (N=13,717), leukocytes (N=91,136), CRP (N=78,085) and haemoglobin (N=92,304).
*Used as reference category in the multivariable logistic regression analyses.
The p-value is presented from the interaction term of age*biomarker which was added in the analyses. Values in bold are statistically significant.

Alternatively, we tested whether age affected the association between biomarkers and mortality using an interaction term. As shown by the significant interaction terms, age significantly affected the associations between urea (p=0.03), creatinine (p<0.01), sodium (p<0.01), hemoglobin (p<0.01), CRP (p<0.01) and lactate (p<0.01) and mortality, but not between leukocytes and mortality (p=0.47). Table 2 shows in which blood test categories the risk on mortality was affected by age.

Supplementary file 5 shows the associations between biomarker categories and mortality in all four age categories, corrected for potential confounders like disease severity and complexity. The results are comparable with the main analyses shown in figure 1 and 2, with larger absolute increases of mortality in older patients with abnormal biomarker values.

For objective 2, figure 3 shows that the risk for absolute case-mix adjusted mortality had larger increases in older patients with an increasing number of abnormal blood test values. Older patients more often had abnormal blood test values (see supplementary file 6).

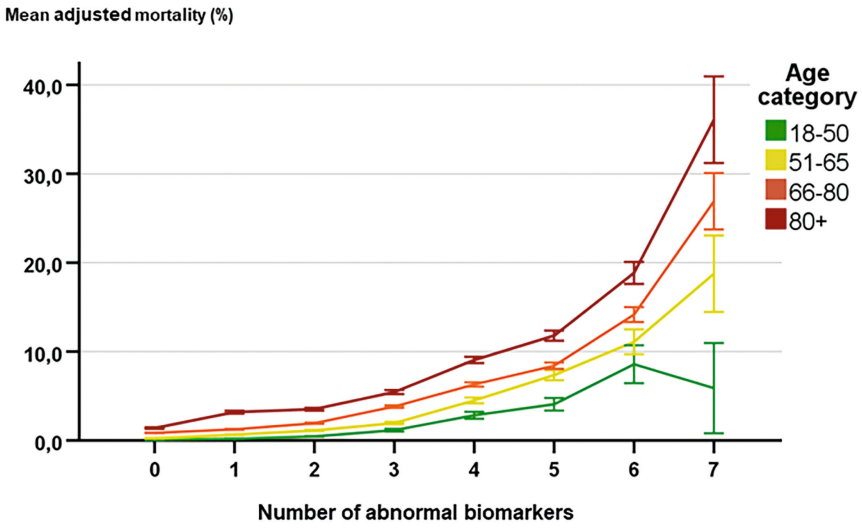


Figure 3 The association between the number of abnormal biomarkers (outside of the commonly used reference ranges) and mortality in different age categories. Mortality is shown as predicted mortality in percent (between 0 and 100%).

DISCUSSION

The present study has two main findings. First, mortality risk for ED patients deteriorates gradually or in a U-shaped fashion with most blood tests. Secondly, the association between blood tests and (adjusted) mortality depends on age, whereas mortality risk is affected most by deviating biomarker levels in younger patients. Age should therefore be considered as an effect modifier rather than a covariate in future prediction models.

In ED-patients, case-mix adjusted mortality increased gradually with increasing urea, CRP, and lactate values, and with the number of abnormal blood tests. For creatinine, sodium, leukocytes, and hemoglobin we found U-shaped associations. These findings correspond to previous studies in which blood tests had similar unadjusted associations with relevant clinical outcomes.[65, 66] Several studies used biomarkers in prediction models for the ED,[67-75] with similar associations for mortality. However, to the best of our knowledge, none of these studies investigated the effect of age on the association between biomarkers and outcome. As we showed, the associations between blood tests and case-mix adjusted outcomes changed with age, suggesting that age should be considered as an effect modifier rather than a covariate, possibly because reference intervals of biomarkers change with age. For example, sodium and hemoglobin levels decline with increasing age in a healthy population,[76, 77] which may explain our findings that low hemoglobin and low sodium have higher odds on mortality in younger compared to older patients. For urea and creatinine, the AORs for mortality were highest in younger patients, which is in accordance with higher reference values for creatinine and urea in older people.[62] Although younger patients, compared with older patients, had more marked increases in AORs for mortality with deviating urea, creatinine, sodium, CRP, hemoglobin, and lactate, the absolute mortality increase was highest in older patients. This can be explained by the higher overall risk for mortality in older patients, irrespective of the value of plasma biomarker levels.

We also showed that abnormal biomarkers are more often present in older patients. In addition, case-mix adjusted mortality had larger absolute increases in older compared to younger patients with an increasing number of abnormal biomarker values, which has not been studied before. The number of abnormal biomarkers had a gradually increasing association with mortality as described before.[78]

Our results imply that risk stratification tools and acute care guidelines for the ED could be improved by incorporating the absolute risk per age

category. For example, in current practice, using a biomarker score based on abnormal blood test values, both younger and older patients may be considered as low risk with two abnormal blood tests. However, a mortality risk increase of 100% compared to baseline risk, may result in for example 10% mortality in older patients (baseline risk 5%) and only 2% mortality in younger patients (baseline risk 1%). As a result, older patients may be considered as high risk with similar biomarker values while younger patients are at low or intermediate risk. By using different risk scores for age categories based on absolute measures of mortality, risk stratification may be improved which may lead to better recognition of disease severity, better disposition decisions and consequently lower mortality.

Another finding from our study is that the association between biomarker levels and mortality is gradual with a linear or U-shaped relation. Consequently, risk stratification can be improved by using numerical scores rather than a single cut-off for every biomarker,[21, 22, 39, 56, 58, 59] as is currently done in the ICU with the Acute Physiology And Chronic Health Evaluation (APACHE) models, in which more points are assigned if blood tests deviate more from the reference range.[54] Although a single cut-off for each biomarker is commonly used in current risk stratification tools, such as the CURB-65 or the SIRS,[21, 22, 39, 56, 58, 59] our findings suggest not to use a single cut-off for any blood test.

Although this study has its strengths, like the large sample size in multiple EDs, there are also limitations. First, the NEED lacks information about comorbidities. Instead, we had to use proxies known to be associated with comorbidities and complexity to overcome this [45]. Second, inherent to retrospective studies, our study could have been subjected to documentation errors, although this was largely minimized by automatization. Finally, we had no follow-up data, so we had to assume that discharged patients did not experience the outcome.

In summary, the prognostic value of blood tests changes with age and mortality risk deteriorates gradually or in a U-shaped fashion with most blood tests for ED patients. Mortality risk is affected mostly by deviating biomarker levels in younger patients. However, due to higher baseline risk in older patients, small changes in biomarker values lead to high mortality. Furthermore, age affected the association between the number of abnormal blood tests and mortality. Our results have implications for interpretation of biomarkers and their use in existing risk stratification tools and guidelines in the ED. Future studies should develop risk stratification tools for ED patients using age-adjusted numerical scores based on absolute risks.

SUPPLEMENTARY FILE 1

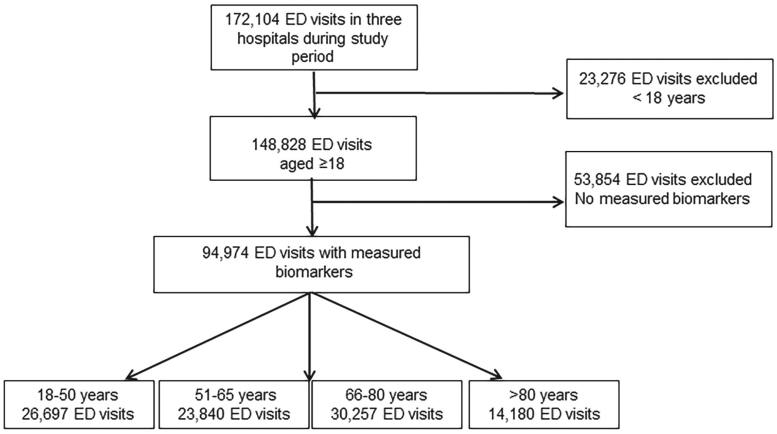
The reference values provided by the hospital laboratories for all studied blood tests, and the type of kit/device that is used.

Blood test	Reference values	Device	Kit
Urea	2.5-7.5mmol/L (men) 2.5-6.4 mmol/L (women)	Cobas 8000	Roche
Creatinine	60-110 µmol/L (men) 49-90 µmol/L (women)	Cobas 8000	Roche
Sodium	135-145 mmol/L	Cobas 8000	Roche
Lactate	0.5-2.2 mmol/L	ABL90	Radiometer
Haemoglobin	8.5-11.0mmol/L (men) 7.5-10.0 mmol/L (women)	XN9100 / Cell-Dyn Sapphire	Sysmex, Abbott
Leukocytes	4.0-10.0 x 10^9/L	XN9100 / Cell-Dyn Sapphire	Sysmex, Abbott
C-reactive Protein	<5.0 mg/L/ <6.0mg/L	Cobas 8000	Roche

SUPPLEMENTARY FILE 2

See supplementary file 2 of Chapter 2 in this thesis.

SUPPLEMENTARY FILE 3



Patient flow diagram of the study.

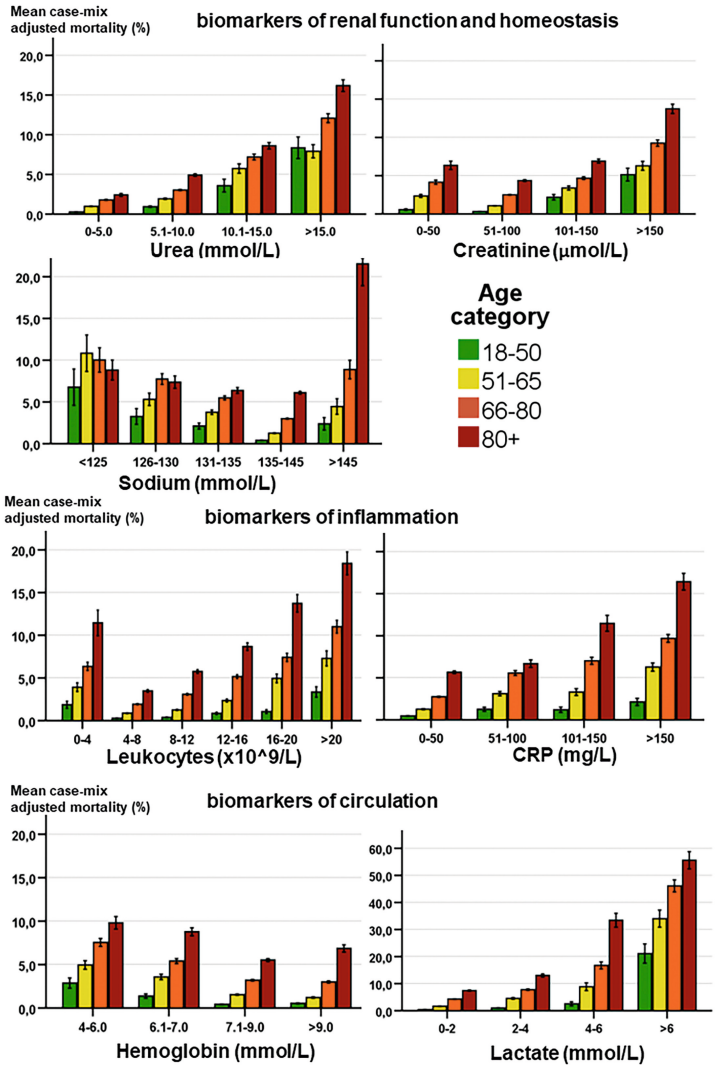
SUPPLEMENTARY FILE 4

Disposition and outcome of included ED patients per age category in whom biomarkers were assessed.

	Total cohort N=94,974	**18-50 years** N= 26,697	**51-65 years** N=23,840	**66-80 years** N=30,257	**>80 years** N=14,180
Patient disposition N (%)					
Discharged home	33,711 (35.5)	13,412 (50.2)	8901 (37.3)	8408 (27.8)	2990 (21.1)
Discharged outpatient clinic	4121 (4.3)	1554 (5.8)	1032 (4.4)	1113 (3.7)	422 (3.0)
Admission to ward	48,849 (51.4)	10,093 (38.2)	11,711 (49.5)	17,651 (58.3)	9394 (66.2)
Admission to MCU/ CCU	3407 (3.6)	423 (1.6)	984 (4.2)	1403 (4.6)	597 (4.2)
Admission to ICU	1799 (1.9)	444 (1.7)	507 (2.1)	679 (2.2)	169 (1.2)
Transfer to a different hospital	2103 (2.2)	405 (1.5)	448 (1.9)	740 (2.4)	510 (3.6)
Died in the ED	185 (0.2)	24 (0.1)	41 (0.2)	75 (0.2)	45 (0.3)
Missing	741 (0.01)	510 (0.01)		179 (0.01)	52 (<0.01)
Outcome N (%)					
In-hospital mortality	2550 (2.7)	153 (0.6)	421 (1.8)	1086 (3.6)	890 (6.3)

Abbreviations: N= number, ED= Emergency Department, CCU = Coronary Care Unit, MCU = Medium Care Unit, ICU = Intensive Care Unit, LOS =Length of Stay, IQR = Inter Quartile Range.

SUPPLEMENTARY FILE 5



Associations between biomarker categories and case-mix adjusted mortality (presented as predicted mortality risk in percentages) is shown for four age categories. The associations are adjusted for sex, age, biomarkers, vitals, proxies for disease severity, complexity and comorbidities.

SUPPLEMENTARY FILE 6

Number of abnormal blood test values per age category

	18-50 years	51-65 years	66-80 years	>80 years
Number of abnormal blood test values N (%)				
0	7478 (28.0)	5368 (22.5)	4868 (16.1)	1410 (9.9)
1	8258 (30.9)	5882 (24.7)	5799 (19.2)	2200 (15.5)
2	6026 (22.6)	5200 (21.8)	6194 (20.5)	2800 (19.7)
3	2949 (11.0)	3569 (15.0)	5497 (18.2)	2974 (21.0)
4	1256 (4.7)	2171 (9.1)	4273 (14.1)	2473 (17.4)
5	535 (2.0)	1138 (4.8)	2516 (8.3)	1631 (11.5)
6	128 (0.7)	448 (1.9)	960 (3.2)	610 (4.3)
7	17 (0.1)	64 (0.3)	150 (0.5)	82 (0.6)
Abnormal Creatinine, N (%)	4048 (16.4)	5692 (25.1)	9498 (33.0)	6035 (44.3)
Abnormal Urea, N (%)	3852 (15.7)	6096 (27.0)	12,677 (44.5)	8544 (64.8)
Abnormal Sodium, N (%)	2314 (9.2)	3731 (16.1)	6255 (21.2)	3525 (25.5)

Number of abnormal blood test values per age category (continued)

	18-50 years	51-65 years	66-80 years	>80 years
Number of abnormal blood test values N (%)				
Abnormal CRP, N (%)	8880 (63.6)	9852 (74.1)	13,888 (77.5)	6599 (81.4)
Abnormal leukocytes, N (%)	11,268 (44.4)	9862 (43.0)	12,635 (43.4)	5924 (43.2)
Abnormal Hemoglobin, N (%)	5405 (21.1)	7051 (30.4)	12,298 (41.6)	6771 (48.8)
Abnormal Lactate, N (%)	1522 (33.2)	1709 (34.1)	2665 (35.0)	1370 (36.3)

Number of patients in the total cohort was for creatinine (N=89,784), urea (N=88,816), sodium (N=91,617) and lactate (N=13,717), leukocytes (N=91,136), CRP (N=78,085) and haemoglobin (N=92,304).
The percentage of patients with abnormal biomarkers, values outside of the reference range, are presented in the table. The used reference ranges were: Creatinine (49-90 µmol/L for women, 60-110 µmol/L for men), Urea (2.5-6.4mmol/L for women, 2.5-7.5mmol/L for men), Sodium (136-145mmol/L), CRP (<10mg/L), Leukocytes (4.0-10.0x10^9/L), Haemoglobin (7.5-10.0mmol/L for women, 8.5-11.0mmol/L for men), Lactate (<2.0mmol/L).


EMERGENCY

4

THE DIFFERENCE BETWEEN THE PATIENTS' INITIAL AND PREVIOUSLY MEASURED SYSTOLIC BLOOD PRESSURE AS PREDICTOR OF MORTALITY IN OLDER EMERGENCY DEPARTMENT PATIENTS

Bart GJ Candel
Iris B van Ingen
Iris PH van Doormalen
Wouter Raven
Lisette AA Mignot
Evert de Jonge
Bas de Groot

Published in European Geriatric Medicine – 2022 Volume 13, issue 2, p359-365

ABSTRACT

Purpose To assess how often baseline systolic blood pressure (SBP) could be retrieved from the Electronic Health Record (EHR) in older Emergency Department (ED) patients. Secondly, to assess whether the difference between baseline SBP and initial SBP in the ED (ΔSBP) was associated with mortality.

Methods A multicenter hypothesis-generating cohort-study including patients ≥70years. EHRs were searched for baseline SBPs. The association between ΔSBP and 30-day mortality was investigated.

Results. Baseline SBP was found in 220 out of 300 patients (73.3%; 95%CI 68.1-78.0%). In 72 patients with normal initial SBPs (133-166 mmHg) in the ED, fifteen (20.8%) had a negative ΔSBP with 20.0% mortality. A negative ΔSBP was associated with 30-day mortality (AHR 4.7; 1.7-12.7).

Conclusions. Baseline SBPs are often available in older ED patients. The ΔSBP may have prognostic value and may be used as an extra variable to recognize hypotension. Future studies should clarify whether the ΔSBP improves risk stratification in the ED.

INTRODUCTION

Due to higher prevalence of hypertension in older patients,[79, 80] a seemingly normal blood pressure in older Emergency Department (ED) patients may in fact be relative hypotension and may not be recognized as such, possibly leading to under-treatment and higher mortality.

Therefore, knowledge of an individual ED patient's baseline systolic blood pressure (SBP) may be essential to assess whether the initial SBP reflects hypotension or normotension. Unfortunately, baseline SBP may often not be available in the ED as Electronic Health Record (EHR) are not routinely shared among general practitioners and hospitals,[81] due to privacy laws and under registration.

Therefore, the aim of the present hypothesis-generating study is two-fold: First, to assess how often individual baseline SBPs are available in older ED patients, and secondly, to evaluate whether the absolute reduction of SBP compared to the patients' baseline SBP value (ΔSBP) is associated with 30-day mortality. The prognostic value of ΔSBP has not been studied before using the patients' individual baseline SBP values. If baseline SBPs are often available and the ΔSBP is associated with 30-day mortality, this variable may contribute to better risk stratification leading to better recognition of vital threads in ED patients.

METHODS

Study design and setting

This multicenter retrospective cohort study was conducted in two EDs in the Netherlands: a tertiary care center with approximately 25,000 ED visits annually and an urban care center with approximately 20,000 ED visits annually. Data were collected from 1 to 27 January 2018. The study was approved by the medical ethics committee of the MMC. The study was registered in the Netherlands Trials Register: NL9029.

Patient selection

During the study period, all consecutive ED patients ≥70 years who were subsequently hospitalized were included (the prevalence of hypertension increases above this age, according to Statistics Netherlands (ww.cbs.nl)). Patients were excluded if non-urgently triaged according to the Manchester or Dutch Triage System or transferred to another hospital.

Outcomes
The primary outcome was the number (N) (%) of patients with an SBP documented (objective 1) or 30-day mortality (objective 2). If patients were discharged alive and 30-day mortality was not known, they were considered alive after 30 days.

Sample size
See supplementary file 1.

Data collection
Data were collected from the EHR. One researcher in each hospital (IvI, IvD) extracted the data. No clear definition exists in the literature of a baseline SBP. In our study, the baseline SBP was defined as the mean of a maximum of the three most recent SBP values from the General Practitioner (GP) referral letters and the three most recent SBP values from the outpatient clinical records. SBP values were not considered as a baseline value if the notes or records described signs for acute illness (i.e., pneumonia, urinary tract infection or fever) or acute pain (i.e., trauma, acute abdominal pain), or if patients were referred to the ED on the same day. The ΔSBP was calculated by subtracting the baseline SBP from the initial SBP in the ED. Thus, a negative ΔSBP indicated that the SBP in the ED was lower than the patients' baseline SBP.

Statistical analyses
Patient characteristics were described for survivors and non-survivors as mean (standard deviation (SD)), median (interquartile range (IQR)) or number (N, (%)) as appropriate. Paired Student's t-tests, Wilcoxon Signed rank test and Pearson $\chi 2$ tests were used to compare patient characteristics. The ΔSBP and initial SBP in the ED were categorized in tertiles based on their distribution. The categorized ΔSBP could be negative, normal, and positive. The categorized initial SBP in the ED was considered low, normal, and high. A cox regression analysis was used to assess the crude and adjusted association between ΔSBP, the initial SBP in the ED, and 30-day mortality, adjusted for age, sex, and triage category (urgent or very urgent/immediate) as a measure of disease severity. The Kaplan-Meier method was employed to generate survival plots, using the log-rank test, for 30-day mortality.

A sensitivity analysis was performed limited to patients who had baseline SBP values documented within 5 years prior to ED-presentation, to limit the chance that antihypertensive drugs were prescribed and because individual baseline values may have changed by ageing.[12]

Adjusted Hazard Ratio's (AHRs) are presented with 95% confidence intervals (95% CIs).

An α of 0.05 was considered as statistically significant. Data were analyzed using SPSS (version 25.0, IBM, New York, USA).

RESULTS

During the study period, 3102 patients presented to the ED of which 300 patients met the inclusion criteria (figure 1). Baseline SBP values were found in 220 patients (73.3%; 95% CI 68.1-78.0%). In the urban care center, a baseline SBP was found in 117 out of 141 patients (83.0%; 75.7-88.8%) compared to 103 out of 159 patients (64.8%; 56.8-72.2%) in the tertiary care center (difference: 18.2%; 8.5-27.9%). 30-day mortality was 14.3% (10.3-19.6%). Differences in patient characteristics existed between survivors and non-survivors (see table 1).

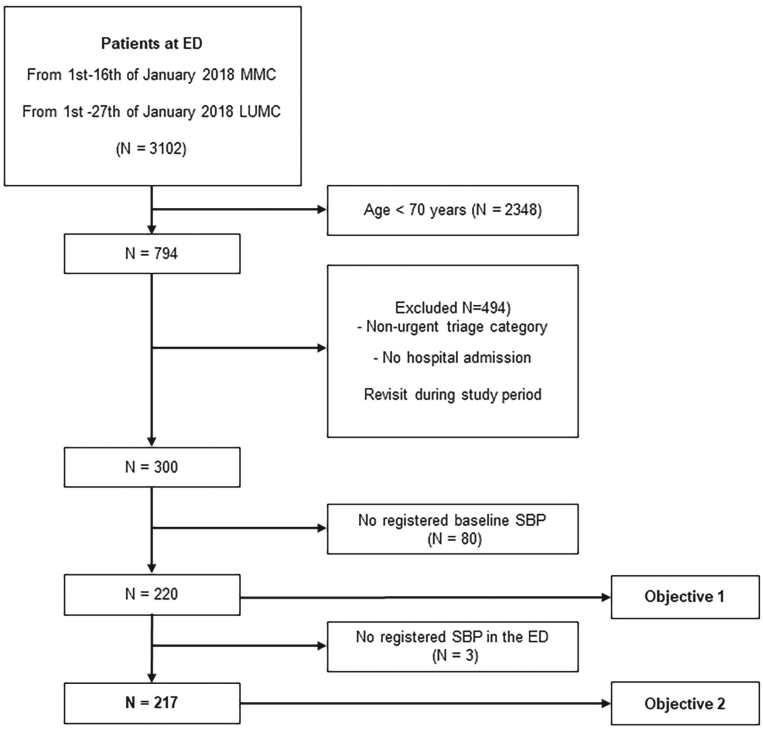


Figure 1 Patient flow and inclusion through the study. The Electronical Patient Records were used to retrospectively include patients in two hospitals starting at 1st January 2018.

Table 1 Patients characteristics

	Total cohort N=217 (100%)	**Survivors** N = 196 (90.3%)	**Non-survivors** N = 21 (9.7%)	**P value**
Demographics				
Age, years, mean (SD)	79 (6.6)	79 (6.4)	82 (7.6)	0.04
Females, N (%)	108 (49.8)	100 (51.0)	8 (38.1)	0.36
Tertiary care center, N (%)	100 (46.1)	90 (45.9)	10 (47.6)	1.0
Baseline SBP †				
Baseline SBP, mmHg, mean (SD)	142 (16.8)	142 (16.4)	143 (20.9)	0.74
Months till ED visit, median [IQR]	33 [16 – 56]	32 [16 – 53]	40 [18 – 77]	0.27
Δ SBP, N (%), ‡				0.01
Negative (< -6.5 mmHg)	73 (33.6)	59 (30.1)	13 (61.9)	
Normal (6.5 – 21.7mmHg)	72 (33.2)	70 (35.7)	3 (14.3)	
Positive (>21.7mmHg)	72 (33.2)	67 (34.2)	5 (23.8)	
Comorbidity				
CCI, mean (SD)	5 (3.1)	5 (3.1)	5 (2.9)	0.87
Use of antihypertensive agents, N (%)	161 (74.2)	145 (74.0)	16 (76.2)	0.88
Arrival by ambulance, N (%)	127 (58.5)	112 (57.1)	15 (71.4)	0.26
Triage category				<0.01
Urgent, N (%)	128 (59.0)	121 (61.7)	7 (33.3)	
Very-urgent, N (%)	76 (35.0)	67 (34.2)	9 (42.9)	
Immediate, N (%)	9 (4.1)	4 (2.0)	5 (23.8)	

Table 1 Patients characteristics (continued)

	Total cohort N=217 (100%)	**Survivors** N = 196 (90.3%)	**Non-survivors** N = 21 (9.7%)	**P value**
Specialty				0.48
Internal medicine, N (%)	78 (35.9)	68 (34.7)	10 (47.6)	
Pulmonary medicine, N (%)	53 (24.4)	49 (25.0)	4 (19.0)	
Neurology, N(%)	40 (18.4)	39 (19.9)	1 (4.8)	
Cardiology, N (%)	16 (7.4)	13 (6.6)	3 (14.3)	
Other, N (%)	29 (13.4)	26 (13.3)	3 (14.3)	
Top three chief complaints				0.87
Feeling unwell, N (%)	55 (25.3)	51 (26.0)	4 (19.0)	
Dyspnea, N (%)	50 (23.0)	41 (20.9)	9 (42.9)	
Abdominal pain, N (%)	14 (6.5)	13 (6.6)	1 (4.8)	
Other, N (%)	54 (24.9)	50 (25.5)	4 (19.0)	
Initial vital signs in the ED				
SBP, mmHg, mean (SD)	150 (35.4)	151 (34.8)	137 (39.8)	0.07
SBP, categorized, mmHg, N (%)				0.03
Low (< 133 mmHg)	76 (35.0)	63 (32.1)	13 (61.9)	
Normal (133-166 mmHg)	71 (32.7)	67 (34.2)	4 (19.0)	
High (>166 mmHg)	70 (32.3)	66 (33.7)	4 (19.0)	
HR bpm, median [IQR]	86 [75-104]	85 [75-103]	98 [72-107]	0.27
RR /min, median [IQR]	20 [16-28]	18 [16-25]	28 [22 – 32]	<0.01
SPO2, %, median [IQR]	96 [94-97]	96 [94-97]	94 [90-97]	0.03

Table 1 Patients characteristics (continued)

	Total cohort N=217 (100%)	**Survivors** N = 196 (90.3%)	**Non-survivors** N = 21 (9.7%)	**P value**
O2 administration, N (%)	102 (47.0)	83 (42.3)	19 (90.5)	<0.01
O2, L/min, median [IQR]	0.0 [0.0 – 4.0]	0.0 [0.0-3.0]	4.0 [0.0-6.0]	<0.01
Body temperature, °C, median, [IQR] (26)	37.3 [36.6-38.4]	37.2 [36.7-38.4]	37.5 [36.0-38.4]	0.89
NRS 1-10, median [IQR]	1 [0-3]	1 [0 -3]	0 [0-2]	0.07
Administered fluid in ED, N (%)	75 (34.6)	64 (32.7)	11 (52.4)	0.08
Disposition				<0.01
Admission to a ward, N (%)	171 (78.8)	158 (80.6)	13 (61.9)	
Admission to MCU, N (%)	5 (2.3)	5 (2.6)	0 (0)	
Admission to ICU, N (%)	10 (4.6)	4 (2.0)	6 (28.6)	
Admission to CCU, N (%)	9 (4.1)	8 (4.1)	1 (4.8)	
Admission to Neurocare, N (%)	22 (10.1)	21 (10.7)	1 (4.8)	

Table 1 Patients characteristics (continued)

	Total cohort N=217 (100%)	**Survivors** N = 196 (90.3%)	**Non-survivors** N = 21 (9.7%)	**P value**
Length of hospital stay days, median [IQR]	5 [2-9]	5 [2-9]	3 [2-9]	0.44

Abbreviations: N = number, SD = Standard deviation, IQR = Interquartile range, CCI = Charlson Comorbidity Index, ED = emergency department, SBP = systolic blood pressure, mmHg = millimetre mercury, Δ = delta, DBP = diastolic blood pressure, HR = heart rate, RR = respiratory rate, SPO2 = peripheral oxygen saturation, O2 = oxygen, °C = degree Celsius, NRS (0-10) = numeric rating scale for pain, MCU = medium care unit, ICU = intensive care Unit, CCU = cardiac care unit, Neurocare = high dependency care unit for neurology patients.

†The baseline SBP was defined as a mean of SBP values which included a maximum of the three most recent SBP values from referral letters of the General Practitioner and the three most recent SBP values from the outpatient clinical records. SBP values were not considered as a baseline value if the letters or records described signs for acute illness (i.e. pneumonia, urinary tract infection or fever) or acute pain (i.e. trauma, acute abdominal pain), or if patients were referred to the ED at the same day. The mean time from the measured baseline SBP till ED visit is presented.

‡ The ΔSBP was calculated by subtracting the baseline SBP from the first measured SBP in the ED. A negative ΔSBP therefore indicated that the SBP in the ED was lower than the patients' baseline SBP. the ΔSBP was categorized in tertiles based on distribution.

Missing values (survivors; non-survivors) = Use of antihypertensive agents (2;0), Arrival by ambulance (5;0), Triage category (4;0), responsible specialism (1;0), Presenting complaints (41;3), HR (8;1), RR (34;2), SPO2 (16;2), O2 administration (59;6), Body temperature (23;3), NRS (51;7), Administered fluid in Ed (4;0), Admission to (3;1).

A cross table between the categorized ΔSBP and the initial SBP in the ED showed that 15 out of 72 patients (20.8%; 13.1-31.6%) with a negative ΔSBP had a normal (mid tertile) initial SBP in the ED (see table 2). Mortality was highest if the ΔSBP was negative and the initial SBP in the ED was low, with 15 out of 57 deceased patients (%26.3).

Table 3 shows the crude and adjusted association for the ΔSBP and the initial SBP in the ED and 30-day mortality. Kaplan Meier survival curves confirm the prognostic role of ΔSBP in ED patients ≥70 years (figure 2).

The sensitivity analysis showed similar results (supplementary file 2).

Table 2. A cross table of categorized ΔSBP and the initial SBP in the ED with numbers of 30-day mortality

	Initial SBP in the ED			
	Low, <133mmHg, (N)	**Normal, 133-166mmHg, (N)**	**High, >166mmHg, (N)**	**Total (N)**
Δ SBP (categorized) Negative (< -6.5 mmHg)	57 † 15 (26.3%)	15 † 3 (20.0%)	0 † 0 (0.0%)	72
Normal (-6.5 – 21.7mmHg)	18 † 2 (11.1%)	43 † 2 (4.7%)	12 † 1 (8.3%)	73
Positive (>21.7mmHg)	1 † 0 (0.0%)	13 † 2 (15.4%)	58 † 6 (10.3%)	72
Total (N)	76	71	70	217

SBP = Systolic Blood Pressure, ED = Emergency Department, Δ = delta. The ΔSBP and the initial SBP in the ED were categorized in tertiles based on their distribution. † number of patients who died in 30days, with percentages in that group.

Table 3. The association between a delta systolic blood pressure and 30-day30-day mortality in hospitalized Emergency Department patients ≥70 years.

	30-day mortality			
	Univariable HR (95% CI)	**P-value**	**Multivariable AHR (95% CI)**	**P-value**
ΔSBP (categorized)				
Negative (< -6.5 mmHg)	4.2 (1.5-11.2)	0.01	4.7 (1.7-12.7)	<0.01
Normal (-6.5 – 21.7mmHg)	1.0		1.0	
Positive (>21.7mmHg)	1.7 (0.55-5.1)	0.37	2.0 (0.63-6.1)	0.25
Initial SBP in the ED (categorized)				
Low (<133mmHg)	2.5 (1.0-5.9)	0.05	2.5 (1.0-6.0)	0.04
Normal (133-166mmHg)	1.0		1.0	
High (>166mmHg)	1.0 (0.35-2.9)	0.99	1.1 (0.38-3.1)	0.90

Abbreviations: HR = Hazard Ratio, AHR = Adjusted Hazard Ratio, 95% CI = 95% Confidence Interval, SBP = Systolic Blood Pressure, ED = Emergency Department, Δ = delta.

The association between ΔSBP and 30-day mortality, and initial SBP in the ED and 30-day mortality, were both adjusted for age, sex, and triage category (urgent, or very urgent/immediate). Because of multi-collinearity, we did not include the initial SBP in the model with delta SBP and mortality.

The ΔSBP was calculated by subtracting the baseline SBP from the first measured SBP in the ED. A negative ΔSBP, therefore, indicated that the SBP in the ED was lower than the patients' baseline SBP. The ΔSBP and the initial SBP in the ED were categorized in tertiles based on their distribution.

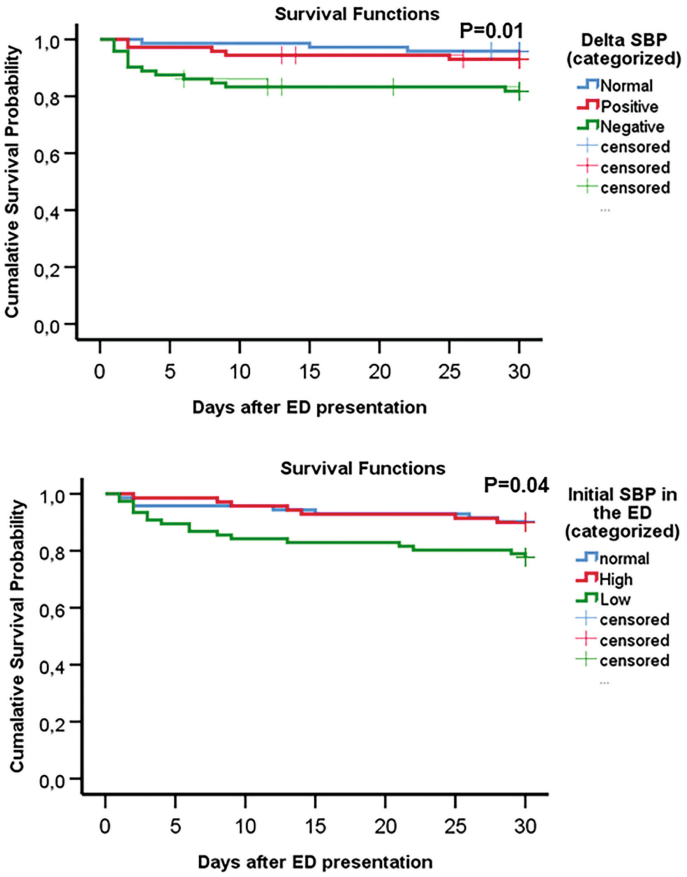


Figure 2 A Kaplan-Meier curve is shown for categorized delta systolic blood pressure and 30-day mortality, and for the initial systolic blood pressure in the Emergency Department and 30-day mortality.

DISCUSSION

In this hypothesis-generating study a baseline SBP value could be retrieved from the EHR in most hospitalized ED patients ≥70 years. In addition, 21% of patients with a normal SBP at ED presentation had a negative ΔSBP and these patients had an increased risk for 30-day mortality.

Low SBP is a well-known risk factor for mortality in older ED patients.[11, 66] Due to the high prevalence of hypertension in the elderly,[79, 82] a seemingly normal SBP may, in fact, be relative hypotension. Consequently, this may lead to an underestimation of the risk and a lower urgency to initiate treatment. Thus,

the deviation from a patients' normal SBP may be relevant. This hypothesis is supported by a study in patients undergoing abdominal surgery, in which an individualized, targeted SBP reduced the risk of postoperative organ dysfunction compared to standard management.[83] This study suggests that individualized treatment using the baseline SBP as a target may improve outcomes. In another study in patients with septic shock, a higher target mean arterial pressure resulted in minor kidney injury among patients known with chronic hypertension.[84]

We are not aware of earlier studies on the association between the difference in SBP between baseline values and SBP at presentation in the ED and outcome in acutely ill patients. However, using another definition, ΔSBP showed prognostic value in trauma patients in the ED.[85-87] In these studies, the ΔSBP was calculated as the difference between the initial SBP in the ED and the prehospital SBP on the same day. A declining SBP in the hours before ED admission was associated with increased mortality, as was also shown for hospitalized patients.[88, 89] In our study, we did not attempt to study short-term SBP trends, but we assessed whether a lower SBP in the ED compared to a baseline SBP was associated with 30-day mortality in older patients. Although baseline SBP in our study could have been measured years before the ED visit, we hypothesized that this SBP would provide important information about the patients' normal value.

Our study was not designed to study whether the ΔSBP was a better predictor for mortality than the initial SBP in the ED. Nonetheless, our study may have implications for clinical practice. We found that even if SBP at admission is normal (> 133 mmHg), some patients had relative hypotension. Future studies should assess whether using the ΔSBP in risk stratification tools like, i.e., the National Early Warning Score has added value to the SBP alone. Also, many challenges are reported regarding health information exchange among hospitals and primary care centers.[81, 90] Because baseline SBPs may have significant prognostic value, our findings argue in favor of promoting health information exchange among hospitals and primary care centers to have baseline vital signs available in the ED.

Our study has several limitations. First, the baseline SBP was defined as a mean of values found in referral letters of GPs and outpatient clinical records. No information existed on how these SBPs were measured. Also, the median time of the measured baseline SBP till ED visit was substantial. Furthermore, antihypertensive agents could have been prescribed in the meantime, or baseline SBPs could have been changed by (patho)physiological changes.[12] We tried to overcome this with a sensitivity analysis using only baseline

SBPs measured <5 years from ED visit, which showed similar results. In addition, baseline SBP should be unaffected by pain, psychological stress, or acute illness. We cannot rule out the possibility that this was not always documented in the files, which could have influenced baseline SBP. Secondly, 30-day mortality was missing in seven patients who were discharged alive. However, five of these patients had a negative ΔSBP and therefore we do not expect this would have affected our results. Lastly, the limited sample size of this study and the two-center design limit the generalizability of our findings and may lead to overestimation of the effect size. Nonetheless, we believe that the ΔSBP needs further investigation in larger prospective studies because it may improve recognition of hypotension.

SUPPLEMENTARY FILE 1

Sample size estimation: In this hypothesis-generating study, we wanted to include 200 patients for objective 2 in both hospitals. We expected that baseline SBP values would be available in approximately 70% of the patients. For objective 1 we therefore expected to include approximately 285 patients in total. Based on a previous study, the mortality in our patient selection would be approximately 8-10 percent, which would result in 16-20 deceased patients.[91] We could therefore adjust for three potential confounders in the multivariable logistic regression analyses according to the existing rule of thumb.[44]

SUPPLEMENTARY FILE 2

Sensitivity analysis, including hospitalized Emergency Department patients ≥70 years, who had baseline SBP values <5 years old till the ED visit.

	In-hospital mortality			
	Univariable OR (95% CI)	**P-value**	**Multivariable AOR (95% CI)**	**P-value**
ΔSBP (categorized)				
Negative (< -6.5 mmHg)	3.8 (1.0-14.6)	0.06	4.2 (1.0-17.1)	0.05
Normal (-6.5 – 21.7mmHg)	1.00		1.00	
Positive (>21.7mmHg)	1.3 (0.3-6.1)	0.73	1.6 (0.3-7.6)	0.58
Initial SBP in the ED (categorized)				
Low (<133mmHg)	3.5 (0.9-13.2)	0.07	3.2 (0.83-12.7)	0.04
Normal (133-166mmHg)	1.00		1.00	
High (>166mmHg)	0.99 (0.19-5.1)	0.99	1.1 (0.2-5.9)	0.90

Abbreviations: OR = Odds Ratio, AOR = Adjusted Odds Ratio, 95% CI = 95% Confidence Interval, SBP = Systolic Blood Pressure, ED = Emergency Department, Δ = delta.
The sensitivity included 185 patients in total, including 16 non-survivors.
The multivariable regression analysis was adjusted for age, sex and triage category (urgent, or very urgent/immediate).
The baseline SBP contained one SBP value closest to the ED visit and <5 years. The ΔSBP was calculated by subtracting the baseline SBP from the first measured SBP in the ED. A negative ΔSBP therefore indicated that the SBP in the ED was lower than the patients' baseline SBP. The ΔSBP and the initial SBP in the ED were categorized in tertiles based on their distribution.


EMERGENCY

5

SEX DIFFERENCES IN CLINICAL PRESENTATION AND RISK STRATIFICATION IN THE EMERGENCY DEPARTMENT: AN OBSERVATIONAL MULTICENTER COHORT STUDY.

Bart GJ Candel
Saimi Dap
Wouter Raven
Heleen Lameijer
Menno I Gaakeer
Evert de Jonge
Bas de Groot

Published in the European Journal of Internal Medicine- 2022, Volume 95, p74-79

ABSTRACT

Objective The aim of this study was to investigate whether sex differences exist in disease presentations, disease severity and (case-mix adjusted) outcomes in the Emergency Department (ED).

Methods Observational multicenter cohort study using the Netherlands Emergency department Evaluation Database (NEED), including patients ≥18 years of three Dutch EDs. Multivariable logistic regression was used to study the associations between sex and outcome measures in-hospital mortality and Intensive Care Unit/Medium Care Unit (ICU/MCU) admission in ED patients and in subgroups of triage categories and presenting complaints.

Results Of 148,825 patients, 72,554 (48.8%) were females. Patient characteristics at ED presentation and diagnoses (such as pneumonia, cerebral infarction, and fractures) were comparable between sexes at ED presentation. In-hospital mortality was 2.2% in males and 1.7% in females. ICU/MCU admission was 4.7% in males and 3.1% in females. Males had higher unadjusted (OR 1.34(1.25-1.45)) and adjusted (AOR 1.34(1.24-1.46)) risks for mortality, and unadjusted (OR 1.54(1.46-1.63) and adjusted (AOR 1.46(1.37-1.56)) risks for ICU/MCU admission. Males had higher adjusted mortality and ICU/MCU admission for all triage categories, and with almost all presenting complaints except for headache.

Conclusions Although patient characteristics at ED presentation for both sexes are comparable, males are at higher unadjusted and adjusted risk for adverse outcomes. Males have higher risks in all triage categories and with almost all presenting complaints. Future studies should investigate reasons for higher risk in male ED patients.

INTRODUCTION

Sex differences affect patient outcomes in many diseases, such as Coronavirus Disease 2019 (COVID-19), acute myocardial infarction, acute ischemic stroke and infectious diseases, while outcome differences in out of hospital cardiac arrest are still controversial.[14-20] Male sex was found to be an independent predictor of mortality in the Emergency Department (ED),[92-95] but it is unclear whether this higher risk for men is caused by male sex itself, by higher disease severity at ED presentation, or by sex related differences in the recognition and treatment of patients at risk in the ED.

We live in a society where sex-specific medicine acquires more attention and physicians become more aware of sex differences in acute diseases.[96] If healthcare providers are aware of sex differences in disease presentations and clinical outcomes, they may be able to better anticipate the clinical course and provide better treatment choices for both men and women. If sex differences in characteristics at ED presentation and (case-mix adjusted) outcomes exist, risk stratification tools, triage systems, and guidelines for management of patients in the ED may need to be sex-adjusted.[39-41]

The aim of this study is therefore twofold: first, to assess sex differences in ED presentations (such as vital signs, triage categories, presenting complaints, blood values, pain scores), and differences in unadjusted and case-mix adjusted clinical outcomes (in-hospital mortality, hospital, and ICU/ Medium Care Unit (MCU) admissions and ED/hospital Length of Stay (LOS)). Secondly, to assess the relation between sex and outcomes in subgroups of patients by triage-category and presenting complaints.

METHODS

Study design and setting

This observational multicenter cohort study was conducted in three EDs in the Netherlands, with each approximately 25,000-30,000 ED visits per year. In this study an existing database was used in which all ED visits were registered of one tertiary care center (1 January 2017 – 8 June 2019), and two urban hospitals (from 1 January 2019 – 12 January 2020 and in one center from 1 January 2017 – 31 December 2019). The study was approved by the medical ethics committee of the Máxima Medical Centre. The study was registered in the Netherlands Trials Register: NL9030.

Selection of participants

All consecutive ED patients ≥18 years were included if sex was registered.

Data collection
Data were collected from the Netherlands Emergency department Evaluation Database (NEED), the national quality registration for EDs in the Netherlands (see www.stichting.need.nl). Data from three hospitals were available. The tertiary care center used the Manchester Triage System (MTS) and both urban hospitals the Dutch Triage Standard (NTS). Previously, we described in detail which data were collected, and how data were categorized and synchronized among hospitals before analyses was possible.[91] In addition, presenting complaints for both triage systems were merged and a National Early Warning Score (NEWS) was calculated (supplementary file 1).[9]

Outcome measures
The primary outcome was in-hospital mortality. Secondary outcomes were ICU/MCU admissions (including Coronary Care Unit admissions), ED-length of stay (0-4hours or ≥4hours), hospital-length of stay, (0-5 days or ≥5 days; 0-30days or ≥30 days) and total hospital admissions. Patients who required, or were likely to require, advanced respiratory support, inotropic agents or were at high risk for cardiac arrest, were admitted to the ICU/MCU if a specialist or intensivist considered this to be necessary. ICU/MCU admission did not include patients who were first admitted to a general ward and were later admitted to an ICU/MCU. A prolonged ED-length of stay was ≥4hours based on previous research.[97] Hospital-length of stay was dichotomized in two ways: shorter or longer than the median length of stay in the Netherlands (5 days),[98] and shorter or longer than 30 days as measure of prolonged hospital stay.[99] Patients discharged from the ED were not included in this outcome. The total hospital admissions included deceased patients in the ED and patients transferred to other hospitals.

Sample size estimation
We aimed to adjust for 12 potential confounders in the regression analyses, mentioned in the 'Statistical analyses' section. Most variables were categorized and had to be used as dummy variables in the analyses. As a result, we had to include 44 (dummy) variables in the regression analysis. Approximately five to ten events per variable are needed to prevent overfitting in association studies.[44] The NEED contained 148,828 ED visits of patients ≥18 years of age, of whom ~50% (74,000) are women. Estimated in-hospital mortality would be ~3% of the overall population. On average we would have 0.03 x 74,000 = 2220 events per group to adjust for the 44 potential confounders in the analysis with the end-point mortality. This number of events is appropriate to prevent overfitting.

Statistical analyses

For objective 1, descriptive statistics were used to compare ED presentation and outcome between males and females. Data were presented as mean and standard deviation (SD) if normally distributed, skewed data were presented as median and interquartile range (IQR). Categorical data were presented as number and percentage (%). Paired Student's t-tests, Wilcoxon Signed rank test and Pearson $\chi 2$ tests (as appropriate) were used to compare patient characteristics between males and females. NEWS scores were presented with the commonly used trigger levels.[9] A sample of the top five diagnoses (using the International Classification of Diseases 10th Revision (ICD-10 codes)) were presented per sex and top 10 presenting complaints, from two out of three hospitals, as information was not available in the third hospital. Univariable and multivariable logistic regression analyses were performed to assess the association between sex and outcomes. All potential confounders considered relevant by three researchers (BC, SD, and BdG) were used in the multivariable model. Vital signs were categorized in four or five categories based on expected distribution to overcome non-linear associations,[82] including a category 'not measured' to prevent information bias by missing data. Multicollinearity was considered not to be a problem if the Variance Inflation Factor (VIF) was below three. Separate models were made for the two different outcomes. Per outcome measure, two different models were constructed. The first model was built to adjust for reason of presentation in the ED and contained the following potential confounders: age, hospital location and presenting complaints (top ten most frequent presenting complaints + miscellaneous presenting complaints). The second model was also adjusted for severity of illness and treatments given in the ED, and contained 12 potential confounders: age, hospital location, presenting complaints, performed radiological tests (0= none performed, 1= X-ray, ultrasound or CT-scan performed), laboratory test score (0= no blood tests performed, 1= any blood test performed), amount of fluids administered during ED stay (0=no fluid, 1= 0-500ml, 2=>500ml), number of consultations in the ED (0=none, 1=1, 2=2, 3= ≥3 consultations), respiratory rate (not measured, 0-9, 10-19, 20-29, ≥30/min), peripheral oxygen saturation (not measured, 0-80, 81-85, 86-90, 91- 95, 96-100%), systolic blood pressure (not measured, 0-80, 81-100, 101-120, 121-140, >140mmHg), heart rate (not measured, 0-50, 51-75, 76-100, 101- 125, >125beats/min.) and temperature (not measured, 0-30, 31-34, 35-37, 38-39, ≥40°C). Vital signs and fluid administration were used as a measure of disease severity. The number of consultations, radiological tests and laboratory test score were used as a measure of comorbidities/complexity as described previously.[45] Patients with missing data were excluded from the analyses. Crude and adjusted odds ratio's (AORs) were reported with 95% Confidence Intervals (95%-CIs).

For objective 2, we assessed the association between triage categories (1= non-urgent, 2= urgent, 3= very urgent, 4= most urgent), presenting complaints and mortality, and ICU/MCU admission with a new multivariable model for men and women separately, using similar potential confounders as used in model 2 described above. Mean adjusted mortality with 95% CIs for each triage category and presenting complaint were presented, which is the absolute mortality after correction for confounding.

Sensitivity analyses
A sensitivity analysis was performed to assess whether using different potential confounders in the multivariable logistic regression affected the impact of sex on outcome. Therefore, we repeated the multivariable model 2 as described above to assess the association between sex and outcome, but in which all vital signs were replaced as potential confounders by NEWS, and blood tests (urea, hemoglobin, and leukocytes) divided in categories, including a category 'not measured'. If vital signs were not measured, they were considered normal values in the NEWS.

In a second sensitivity analysis we excluded all patients who died in the ED. After exclusion of these patients, we repeated the multivariable logistic regression model 2 and reported the AOR for male sex.

Subgroup analyses
To assess whether etiology affected the association between sex and the primary outcome, subgroup analyses were performed selecting patients with a suspected infection, trauma, and chest pain. Supplementary file 2 describes how these patients were selected in the database.

IBM SPSS Statistics version 25.0 was used for all statistical analyses. A P-value <0.05 was considered significant.

RESULTS

Patient characteristics
The total cohort consisted of 148,825 ED visits, of which 48.8% were women. Supplementary file 3 shows the flow chart of the present study. The mean age for men was 56.0 (20.0) years and for women 56.8 (20.9) years. Only small differences existed between men and women in demographics, characteristics during ED visit, vital signs, and performed diagnostics (table 1). For example, self-referral was higher among men (43.8%) compared to women (37.5%), (difference 95%-CI 5.8-6.8%). Males more often presented with chest pain (8.8%) than women (7.4%) (difference, 95%-CI 1.0-1.6%). The

miscellaneous presenting complaints consisted of 41 presenting complaints (supplementary file 1), including urinary tract problems, eye problems, suicidal behavior, intoxications, diarrhea/vomiting, and fall which all had an incidence of >1.0% without differences between men and women. All other presenting complaints had an incidence <1.0%. Males were more often triaged as 'most urgent' (4.7%) than women (3.7%) (difference, 95%-CI 0.7-1.1%). Severe pain (Numeric Rating Scale ≥7) was more often reported by women (7.6%) compared to men (5.8%) (difference 95%-CI 1.1-2.6%). Creatinine, urea, and hemoglobin levels were higher in male sex. A sample of the top five diagnoses per sex and presenting complaints from two hospitals showed roughly similar results for men and women (supplementary file 4). Patient characteristics were also described separately for patients triaged as 'very urgent' and 'most urgent', which showed similar results (supplementary file 5).

Table 1. Patient characteristics

	Total cohort	Male patients	Female patients
Demographics			
N (%)	148,825	76,271 (51.2)	72,554 (48.8)
Age, years, mean (SD)	56.4 (20.5)	56.0 (20.0)	56.8 (20.9)
Tertiary care center	55,069 (37.0)	28,786 (37.7)	28,283 (36.2)
Arrived by ambulance, N (%)	45,148 (33.1)	23,425 (33.4)	21,723 (32.7)
Self-referral, N (%)	59,718 (40.1)	32,811 (43.8)	26,907 (37.5)
Trauma/shock room, N (%)	9,689 (10.5)	5,040 (10.8)	4,649 (10.2)
Characteristics during ED-visit			
Top 10 presenting complaints, N (%			
1. Extremity problems	26,981 (18.1)	13,003 (17.0)	13,978 (19.3)
2. Feeling unwell	24,139 (16.2)	12,567 (16.5)	11,572 (15.9)
3. Abdominal pain	16,032 (10.8)	6835 (9.0)	9197 (12.7)
4. Dyspnea	13,049 (8.8)	6457 (8.5)	6592 (9.1)
5. Chest pain	12,077 (8.1)	6675 (8.8)	5402 (7.4)
6. Wounds	7,102 (4.8)	4604 (6.0)	2498 (3.4)
7. Trauma	5,056 (3.4)	2769 (3.6)	2296 (3.2)
8. Collapse	4,385 (2.9)	2542 (3.3)	1843 (2.5)
9. Palpitations	3,746 (2.5)	2045 (2.7)	1701 (2.3)
10. Headache	2,503 (1.7)	1068 (1.4)	1435 (2.0)
Miscellaneous	33,746 (22.7)	17,706 (23.2)	16,040 (22.1)

Table 1. Patient characteristics (continued)

	Total cohort	Male patients	Female patients
Triage category, N (%)			
Non-urgent	43,259 (30.3)	22,388 (30.7)	20.871 (30.0)
Urgent	60,309 (42.3)	29,772 (40.8)	30,537 (43.9)
Very urgent	33,090 (23.2)	17,442 (22.9)	15,648 (22.5)
Most urgent	5970 (4.0)	3396 (4.7)	2574 (3.7)
NEWS, median (IQR)	3 (3-5)	3 (3-5)	3 (3-5)
NEWS trigger level 1 (0-4), N (%)	106,933 (71.9)	54,778 (71.8)	52,155 (71.9)
NEWS trigger level 2 (5-6), N (%)	23,522 (15.8)	11,748 (15.4)	11,774 (16.2)
NEWS trigger level 3 (7-23), N (%)	18,370 (12.3)	9745 (12.8)	8625 (11.9)
Vital signs			
SBP, mmHg, mean (SD)	133 (31)	132 (31)	133 (32)
DBP, mmHg, mean (SD)	83 (17)	84 (17)	81 (17)
MAP, mmHg, mean (SD)	99 (19)	99 (19)	98 (19)
Temperature, °C, median [IQR]	36.9 [36.5-37.4]	37.0 [36.6-37.4]	36.8 [36.4-37.3]
RR, /min, median [IQR]	17 [14-20]	17.0 [14-20]	17.0 [14-20]
SPO2, %, median [IQR]	98 [96-100]	98.0 [96-99]	98.0 [96-100]
HR, /min, mean (SD)	86 (21)	85 (22)	87 (20)
Pain scores, Numeric Rating Scale, N (%)			
Not measured	77,948 (52.4)	39,812 (52.2)	38,136 (52.6)
0	18,979 (12.8)	10,476 (13.7)	8503 (11.7)

Table 1. Patient characteristics (continued)

	Total cohort	Male patients	Female patients
1-3	22,357 (15.0)	12,154(15.9)	10,203 (14.1)
4-6	19,582 (13.2)	9395 (12.3)	10,187 (14.0)
7-10	9959 (6.7)	4434 (5.8)	5525 (7.6)
Diagnostics			
Radiology performed[a], N (%)	83,886 (56.4)	42,682 (56.0)	41,204 (56.9)
Urine sediment, N (%)	25,876 (26.1)	12,227 (24.4)	13,649 (27.9)
Blood gas, N (%)	22,491 (15.1)	11,969 (15.7)	10,522 (14.5)
Blood cultures, N (%)	13,357 (13.8)	7,261 (14.8)	6,096 (12.8)
Laboratory test score[b], N (%)	94,973 (63.9)	47,981 (62.8)	47,082 (64.9)
Number of consultations, N (%)			
None	58,570 (41.4)	29,944 (41.5)	28,626 (41.3)
1	74,140 (52.4)	37,651 (52.2)	38,489 (52.7)
2	7722 (5.5)	4005 (5.5)	3717 (5.4)
≥3	1032 (0.7)	570 (0.8)	462 (0.7)

Table 1. Patient characteristics (continued)

	Total cohort	Male patients	Female patients
Biomarkers			
Sodium, median [IQR]	140 [137-142]	140 (137-142)	140 (137-142)
Potassium, median [IQR]	4.1 [3.8-4.4]	4.2 (3.9-4.5)	4.1 (3.8-4.4)
Creatinine, mmol/L, median [IQR]	77 [63-97]	87 [74-108]	67 [57-82]
Urea, mmol/L, median [IQR]	5.8 [4.3-7.9]	6.3 [4.9-8.5]	5.2 [3.9-7.1]
CRP, mg/L, median [IQR]	10 [4.-47]	12 [5-53]	10 [4-42]
Leucocytes, x10^9/L, median [IQR]	9.1 [7.0-12.0]	9.1 [7.0-12.1]	9.1 [7.0-12.0]
Hb, mmol/L, mean (SD)	8.3 (1.3)	8.5 (1.4)	8.0 (1.1)
Lactate, mmol/L, median [IQR]	1.6 [1.1-2.3]	1.6 [1.1-2.5]	1.5 [1.0-2.2]

Abbreviations= N: number, SD: standard deviation, IQR: inter quartile range, ED: Emergency Department, NEWS: National Early Warning Score, SBP: systolic blood pressure, DBP: diastolic blood pressure, MAP: mean arterial pressure, RR: respiratory rate, HR: heart rate, SPO2: peripheral oxygen saturation, mmHg= millimetre mercury, °C: degrees Celsius, CRP: C-reactive Protein, Hb: haemoglobin.

[a]: If one or more of the following radiological tests were performed: Ultrasound, radiography, and computer-tomography.

[b:] Laboratory test score reflects whether one or more blood tests were performed.

Missing values (male;female)= age (0:0); arrival type (6220;6056) referral status (1321;850) location (0;0) triage category (3273;2924) presenting complaints (3626;2995) trauma room (29499;26773) radiology performed (107;87) consultations (4101;3260) blood gas (52;48) blood culture (27112;24977) laboratory test score (52;48) sodium (29872;27330) Potassium (30625;28349) creatinine (30910;28124) urea (31198;28804) lactate (65083;62751) CRP (37053;33798) Leucocytes (30681;27572) Hb (29712;26805) SBP (31158;28136) DBP (31181;28169) MAP (31160;28153) temperature (34931;31530) HR (33355;30694) RR (40815;39347) SPO2 (30076;26964).

The association between sex and clinical outcomes

In-hospital mortality was 2.2% in males and 1.7% in females (table 2). ICU/MCU admission was 4.7% in males and 3.1% in females. Females were more often discharged home from the ED (49.7%) compared to males (47.3%), however, 7-day revisit to the ED was similar for men and women.

Table 2. Relevant clinical outcomes in male and female ED patients

	Total cohort N: 148,825	Males N=76,271	Females N=72,554	P value
In hospital-mortality[a], N (%)	2851 (1.9)	1663 (2.2)	1188 (1.7)	<0.01
Disposition, N (%)				
Total hospital admissions	62,472 (42.0)	32,797 (43.0)	29,675 (40.9)	<0.01
ICU/MCU admission	5425 (3.9)	3317 (4.7)	2108 (3.1)	<0.01
Ward admission	54,175 (38.9)	27,936 (39.4)	26,239 (38.4)	<0.01
Transfers to other hospitals	2519 (1.8)	1289 (1.8)	1230 (1.8)	0.84
Discharged home	67,472 (48.5)	33,554 (47.3)	33,918 (49.7)	<0.01
ED-LOS, hours, median [IQR]	2.7 [1.7-3.8]	2.6 [1.7-3.7]	2.8 [1.8-3.8]	<0.01
H-LOS, days, median [IQR]	4.0 [2.0-7.0]	4.0 [2.0-7.0]	4.0 [2.0-7.0]	<0.01
7-day revisit to the ED, N (%)	8015 (5.4)	4068 (5.3)	3947 (5.4)	0.79

Abbreviations= N: number, SD: standard deviation, IQR: inter quartile range, ICU: intensive care unit, MCU: medium care unit, ED-LOS: Emergency Department-Length Of Stay, H-LOS: hospital-Length Of Stay.
[a]: in-hospital mortality included patients who died in the ED.
Missing values (male;female)= mortality (1242;914) Disposition (5306;4279) ED-LOS (15;19) H-LOS (28682;27437)

Males had higher unadjusted risk (AOR 1.34; 1.25-1.45) and adjusted risk for in-hospital mortality with both model 1 (AOR 1.36;1.26-1.47) and model 2 (AOR 1.34; 1.24-1.46). Also, for high dependency care unit admission, unadjusted risk (AOR 1.54;1.46-1.63) and adjusted risk with model 1 (AOR 1.47;1.38-1.55) and model 2 (AOR 1.46; 1.37-1.56) were higher in males (see table 3). Males had a lower unadjusted risk (AOR 0.90; 0.88-0.92) and

adjusted risk (model 2) (AOR 0.89; 0.87-0.92) for a long ED length of stay of ≥4hours.

Table 3. The risk for clinical outcomes for male ED patients compared to female ED patients**.**

	Crude OR (95%-CI)	Model 1 AOR (95%-CI)	Model 2 AOR (95%-CI)
In-hospital mortality	1.34 (1.25-145)	1.36 (1.26-1.47)	1.34 (1.24-1.46)
ICU/MCU admission	1.54 (1.46-1.63)	1.47 (1.38-1.55)	1.46 (1.37-1.56)
Total hospital admissions	1.09 (1.07-1.11)	1.16 (1.14-1.19)	1.18 (1.15-1.21)
ED-LOS (≥4h)	0.90 (0.88-0.92)	0.92 (0.90-0.94)	0.89 (0.87-0.92)
H-LOS (≥5d)	0.96 (0.93-1.00)	0.98 (0.95-1.02)	0.97 (0.94-1.00)
H-LOS (≥30d)	1.14 (1.01-1.30)	1.14 (1.00-1.29)	1.11 (0.98-1.26)

Abbreviations= OR: odds ratio, AOR: adjusted odds ratio, CI: confidence interval, MCU: medium care unit, ICU: intensive care unit, ED-LOS: emergency department-Length Of Stay, H-LOS: hospital-Length Of Stay.
For H-LOS, only patients who were admitted were included in the analyses.
ORs >1 means a higher risk on the outcome for males compared to females.
Model 1 is adjusted for age, top 10 presenting complaints, and hospital location.
Model 2 is adjusted for age, top 10 presenting complaints, hospital location, performed diagnostics (radiology, blood tests), fluid administration, number of consultations and vital signs.

The association between triage categories, presenting complaints, and outcomes.

Men had higher risks on mortality or ICU/MCU admission in all triage categories (Figure 1). In the 'non-urgent' triage-category, adjusted mortality was 0.52% (0.51-0.53%) in men and 0.46% (0.45-0.47%) for women. Adjusted ICU/MCU admission was 0.74% (0.72-0.76%) for men and 0.57% (0.56-0.59%) for women. In the 'most urgent' triage category, adjusted mortality was 16.0% (15.4-16.6%) in men and 12.2% (11.7-12.8%) in women. Adjusted ICU/MCU admission was 23.6% (23.0-24.2%) in men and 18.7% (18.1-19.3%) in women.

With all presenting complaints, except wounds and headache, men had higher adjusted mortality and ICU/MCU admission (see figure 2). Mean adjusted mortality in men and women was 0.34% (0.33-0.35%) and 0.33% (0.32-0.34%) for extremity complaints, 4.5% (4.4-4.6%) and 3.7% (3.6-3.8%) for feeling unwell, 1.6% (1.5-1.7%) and 0.85% (0.81-0.89%) for abdominal pain, 5.6% (5.4-5.8%) and 4.1% (3.9-4.2%) for dyspnea, 1.0% (0.98-1.1%) and 0.70% (0.66-0.73%) for chest pain, 0.090 (0.084-0.093%) and 0.12 (0.11-0.13%) for

wounds, 1.4% (1.3-1.5%) and 1.1% (1.0-1.2%) for trauma, 4.7% (4.5-5.0%) and 3.3% (3.1-3.6%) for collapse, 0.45% (0.41-0.48%) and 0.18% (0.16-0.19%) for palpitations, 0.86% (0.78-0.94%) and 1.9% (1.8-2.1%) for headache and 2.1% (2.0-2.2%) and 1.5% (1.4-1.5%) for miscellaneous.

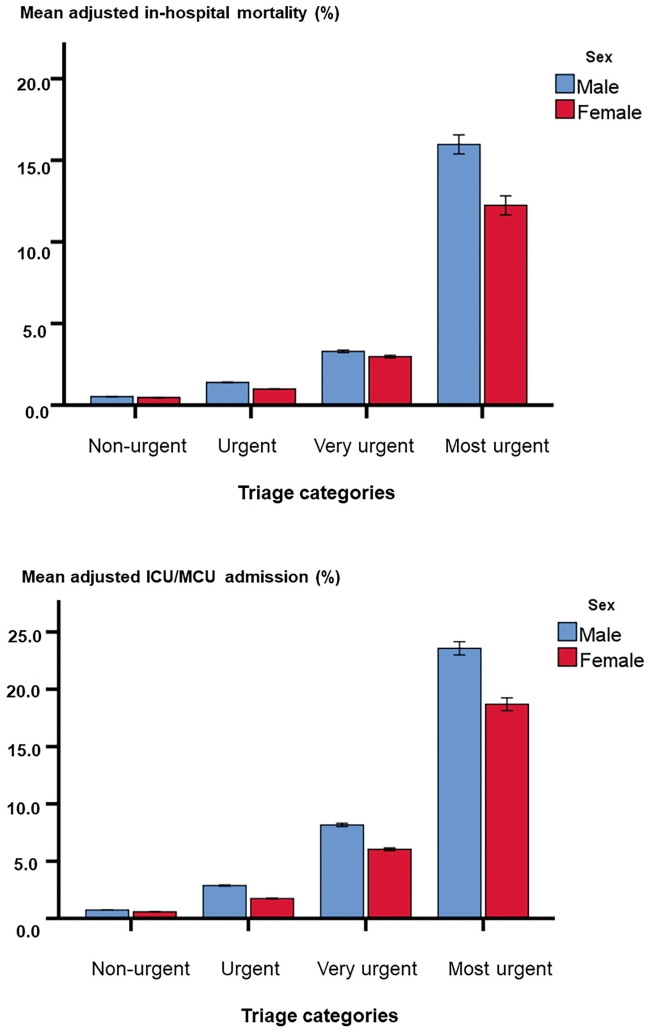
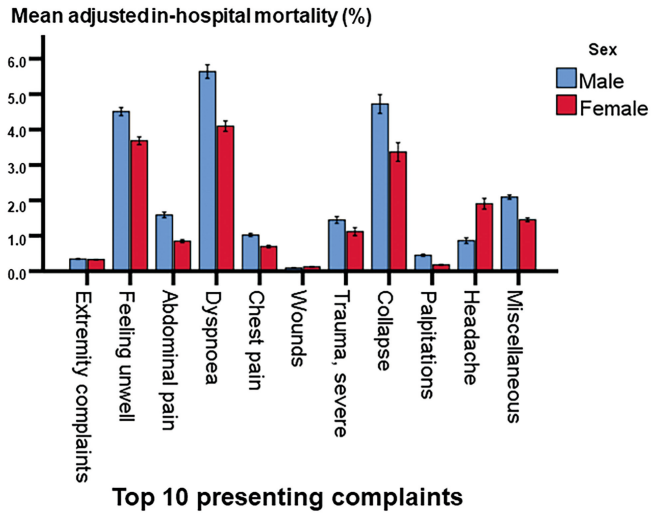


Figure 1 The first panel shows the mean adjusted in-hospital mortality is shown with 95% Confidence intervals for men and women per triage category, according to the Manchester Triage System or the Dutch Triage Standard. The multivariable model was adjusted for age, presenting complaints, disease severity, vital signs, and proxies for comorbidities/complexity. Panel B shows the mean adjusted Intensive Care Unit/Medium Care Unit admission per triage category in men and women.



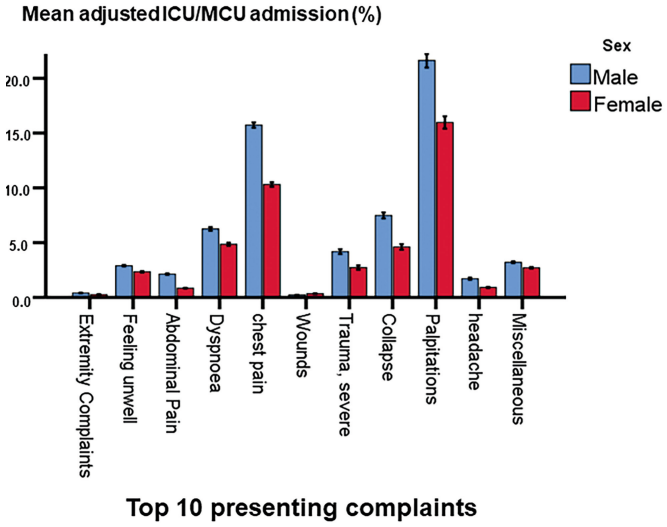


Figure 2 The first panel shows the mean adjusted in-hospital mortality is shown with 95% Confidence intervals for men and women per top 10 presenting complaints, according to the Manchester Triage System or the Dutch Triage Standard. The multivariable model was adjusted for age, presenting complaints, disease severity, vital signs, and proxies for comorbidities/complexity. Panel B shows the mean adjusted Intensive Care Unit/Medium Care Unit admission per top 10 presenting complaints in men and women.

Sensitivity subgroup analyses
The sensitivity analyses showed similar results with higher risks for men if other potential confounders were used for adjustment for severity of illness, and if patients who died in the ED were excluded from the analysis (supplementary file 6). Also, the subgroup analyses for suspected infection, trauma and patients with chest pain showed similar results with higher risks for men (supplementary file 2).

DISCUSSION

Although characteristics at ED arrival are roughly similar, male ED patients have substantially higher crude and adjusted in-hospital mortality and were more often admitted to high dependency care units compared to female ED patients. These increased risks were also present in all triage categories and with almost all presenting complaints (including chest pain), as well as in subgroups of patients with suspected infection and trauma.

To study the nature of the increased risk in males, we first explored whether this could be explained by higher age or different reasons for ED admission. We found that the differences in mortality and admission to high dependency care units persisted after correction for age and presenting complaints. Our results are in line with earlier studies, showing that male sex was found to be an independent predictor of adverse outcomes in patients admitted to an ED with AORs ranging from 1.14 to 1.62.[92, 93, 95] However, in these studies it was not reported whether disease severity in males and females were similar at ED presentation. In our study the increased risk in male patients for adverse outcomes compared with females was still present after adjusting for initial disease severity before ED treatment, suggesting that factors other than age, presenting complaints and initial severity of illness were responsible for higher risk for adverse outcomes.

We can only speculate which other factors may contribute to the increased risk for males compared to females. First, sex itself could be responsible for adverse outcomes in males, meaning that the male phenotype is associated with a more serious course of disease despite similar severity of illness at ED presentation. Differences in immune response, cardiovascular function and sex hormones have been suggested as the possible explanations for better outcomes of females in many diseases, like in sepsis patients.[100] Alternatively, it is possible that male patients present with more serious reasons for admission despite similar presenting complaints and vital signs. Male patients had higher ICU admissions than females, which may illustrate that they were more often non-responders to ED treatment, which is a measure

of disease severity not captured in the initial variables. In addition, similar presenting complaints could represent more severe diseases in males. However, females are more likely to present with nonspecific complaints in for example acute myocardial infarction or stroke.[15, 101] Atypical presentation of disease was associated with adverse outcome.[102] Consequently, one would expect that nonspecific presenting complaints like feeling unwell, palpitations, or dyspnea would most likely represent serious disease in women with consequently higher risk for bad outcomes. In contrast we found lower risks for women, also in these subgroups. Furthermore, we showed that for each presenting complaint, diagnoses were comparable between men and women. Alternatively, we cannot exclude the possibility that differences in presenting complaints between males and females may lead to a different risk estimation in triage systems resulting in delayed initial management in the ED for males.

Differences in ED treatment may be another possible explanation for outcome differences between men and women. Also, males with similar vital signs may be sicker than females due to differences in reference values. For example, a seemingly normal blood pressure in men may actually be relative hypotension, due to higher incidence of hypertension in males, causing end-organ hypoperfusion leading to adverse outcome.[103] If so, risk scores such as the triage system, which are mainly based on vital signs, may have different associations with clinical outcomes for men and women. In our study, in the non-urgent triage category, the absolute risk difference between men and women was only 0.06% while in the most-urgent triage category the difference was 3.8%, which suggests that disease severity is not equally valued in triage systems. Possibly, the same is true for other risk scores such as the NEWS, SIRS or PIRO score. In daily practice, risk stratification tools are used to guide treatment and disposition decisions. For example, in chest pain, the HEART and TIMI scores are used to identify low-risk patients who could be safely discharged from the ED.[104, 105] Studies that evaluated the HEART and TIMI score independently in men and women with chest pain, found a two times higher six-week risk for major adverse cardiac events in men than in women with low risk scores.[105, 106] By using risk stratification tools in the ED, like the HEART or TIMI score, or triage systems, physicians may not recognize potential sex differences in disease severity, which may cause inappropriate disposition decisions (i.e. to a ward instead of to ICU) or wrong initial treatment. Our findings suggest that male sex is an independent risk factor for adverse outcome, and thus, including male sex in scoring systems may improve risk stratification.

Only in patients with headache as chief complaint, women had an increased risk compared to men. It is known that there is an increased risk for subarachnoid hemorrhage in women compared to men, and that women have an increased risk of death due to stroke.[16] Unfortunately, in our dataset, we had no information on incidence of stroke in patients presenting with headache.

Our study has implications for clinical practice. First, sex differences in outcome cannot be explained by differences in age, presenting complaints, or disease severity before ED treatment. Secondly, in future studies, risk stratification tools and acute care guidelines may be improved if sex is implemented as a predictor of adverse outcomes, or if different risk scores are used for men and women, which is currently not done.[39-41, 107]

Despite several strengths like the large sample size and a multicenter design, this study has several limitations. First, the NEED did not contain information about comorbidities/complexity or frailty scores. To overcome this, we used variables that are associated with comorbidities and complexity.[97] Secondly, we had to assume that patients discharged from the ED did not have adverse outcomes as we did not follow-up on them. Nonetheless, we expect this would be similar for both sexes as 7-day ED revisit was also similar. Lastly, the observational nature of our study could have been subjected to errors of documentation and data entry, although this was largely automatized which minimalized accidental misregistration.

In summary, male ED patients had higher risks for crude and adjusted in-hospital mortality and high dependency care unit admission compared to female ED patients, while patient characteristics at ED arrival were roughly similar. Furthermore, in all triage categories and with almost all presenting complaints, men had higher risks than women. Future studies should investigate reasons for higher risk in male ED patients, and whether sex-adjusted risk scores aid to identify patients at risk which may improve outcomes.

SUPPLEMENTARY FILE 1

See also supplementary file 2 of Chapter 2.

To calculate a National Early Warning Score (NEWS), all missing vital signs were considered normal. A Glasgow Coma Scale of 14-15 was considered as Alert from the AVPU score.

Chart for calculation of 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	51-90	91-110	110-130	≥131	
Level of consciousness				A			V, P, or U

A, Alert; V, Verbal; P, Reacting to pain; U, Unconscious; NEWS, National Early Warning Score

SUPPLEMENTARY FILE 2

Subgroup analyses

Suspected infection: Patients with suspected infection were selected if they were hospitalized and blood cultures were obtained in the ED.

Trauma: Trauma patients were selected by triage colour orange or red (urgent) according to the Manchester Triage System or Dutch Triage Standard, and by the following triage presenting complaints: Severe trauma; facial complaints; fall; extremity complaints; large scale incidents; neck pain; thorax injury; head injury; wounds. Additionally, a radiological test had to be performed in the ED, and patients had to be hospitalized or deceased in the ED.

Patients with chest pain: Patients with chest pain were selected based on the chief complaint chest pain, followed by hospitalisation or death in the ED.

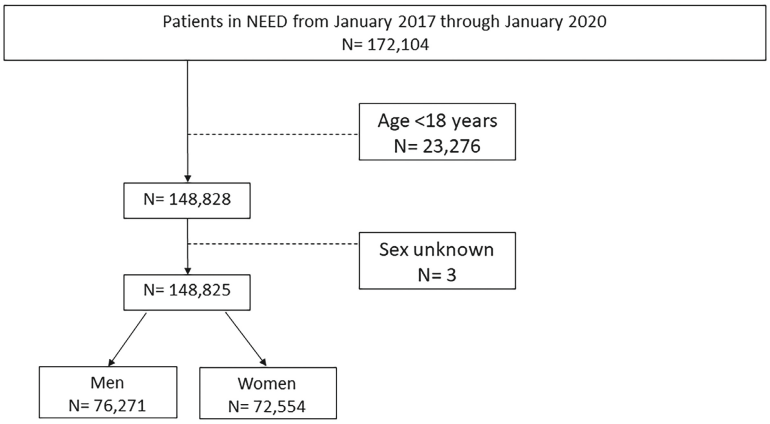
Risk for male ED patients compared to female ED patients on relevant clinical outcomes.

	Crude OR (95%-CI)	Main analysis AOR (95%-CI)
Suspected infection (N= 13,717)		
In-hospital mortality	1.21 (1.05-139)	1.20 (1.03-1.40)
ICU/MCU admission	1.17 (1.00-1.36)	1.17 (0.99-1.39)
Trauma (N=7406)		
In-hospital mortality	1.16 (0.86-1.57)	1.78 (1.29-2.47)
ICU/MCU admission	1.87 (1.47-2.38)	1.50 (1.15-1.95)
Chest pain (N=5220)		
In-hospital mortality	1.13 (0.75-1.70)	1.25 (0.79-1.98)
ICU/MCU admission	1.26 (1.11-1.42)	1.35 (1.19-1.54)

Abbreviations= OR: odds ratio, AOR: adjusted odds ratio, CI: confidence interval, MCU: medium care unit, CCU: ICU: intensive care unit.
Interpretation: OR >1 indicates a higher risk for women compared to men, OR<1 indicates a lower risk for women compared to men.
The main analysis is adjusted for age, top 10 presenting complaints, hospital location, performed diagnostics (radiology, blood tests), fluid administration, number of consultations, vital signs.
ORs >1 means a higher risk on the outcome for males compared to females.

SUPPLEMENTARY FILE 3

Patient flow diagram from the present study.



SUPPLEMENTARY FILE 4

Frequencies and mortality of Top 5 discharge diagnoses (ICD-10 codes) specified by sex and presenting complaints in two out of three hospitals

	Total cohort	n	%	Mortality n(%)	Female	n	%	Mortality n(%)	Male	n	%	Mortality n(%)
Extremity complaints												
No. 1	Fracture wrist/hand	1704	9.7	0(0)	Fracture of forearm	1015	11.1	0(0)	Fracture wrist/hand	911	10.8	0(0)
No. 2	Fracture of forearm	1675	9.5	0(0)	Fracture wrist/hand	793	8.6	0(0)	Fracture of forearm	660	7.8	0(0)
No. 3	Fracture of femur	1190	6.7	27(2.3)	Fracture of femur	784	8.5	14(1.8)	Superficial wounds hand/fingers	548	6.5	0(0)
No. 4	Fracture shoulder/ upper arm	1146	6.5	0(0)	Fracture shoulder/ upper arm	667	7.3	0(0)	Fracture shoulder/ upper arm	479	5.7	0(0)
No. 5	Fracture lower leg/ ankle	1029	5.9	0(0)	Superficial wounds hand/fingers	465	5.1	0(0)	Fracture lower leg/ ankle	477	5.6	0(0)
Feeling unwell												
No. 1	Cerebral infarction	1181	10.7	43(3.6)	Cerebral infarction	556	10.6	22(4.0)	Cerebral infarction	625	10.8	21(3.4)
No. 2	Other urinary diseases	774	7.0	26(3.4)	Other urinary diseases	354	6.7	15(4.2)	Other urinary diseases	420	7.3	11(2.6)
No. 3	Pneumonia	701	6.4	46(6.6)	Pneumonia	300	5.7	17(5.7)	Pneumonia	401	7.0	29(7.2)
No. 4	TIA/cerebrovascular disease	533	4.8	1(0.2)	TIA/cerebrovascular disease	259	4.9	1(0.4)	TIA/cerebrovascular disease	274	4.8	0(0)
No. 5	Malignancy digestive organs	370	3.4	15(4.1)	Metabolic disorders	206	3.9	8(3.9)	Malignancy digestive organs	246	4.3	11(4.5)

Frequencies and mortality of Top 5 discharge diagnoses (ICD-10 codes) specified by sex and presenting complaints in two out of three hospitals (continued)

	Total cohort	n	%	Mortality n(%)	Female	n	%	Mortality n(%)	Male	n	%	Mortality n(%)
Dyspnea												
No. 1	COPD (exacerbation)	1673	20.4	90(5.4)	COPD (exacerbation)	877	21.4	57(6.5)	COPD (exacerbation)	796	19.5	33(4.1)
No. 2	Other heart disease	1380	16.8	98(7.1)	Other heart disease	644	15.7	38(5.9)	Other heart disease	736	18.0	60(8.2)
No. 3	Pneumonia	1309	16.0	109(8.3)	Pneumonia	597	14.5	41(6.9)	Pneumonia	712	17.4	68(9.6)
No. 4	Other respiratory diseases*	867	10.6	58(6.7)	Other respiratory diseases*	409	10.0	27(6.6)	Other respiratory diseases*	458	11.2	31(6.8)
No. 5	Asthma	497	6.1	3(0.6)	Asthma	362	8.8	2(0.6)	Pulmonary embolism	199	4.9	3(1.5)
Abdominal pain												
No. 1	Gallbladder, biliary, pancreas disorders	1386	19.5	10(0.7)	Gallbladder, biliary, pancreas disorders	771	20.3	5(0.6)	Gallbladder, biliary, pancreas disorders	625	18.5	5(0.8)
No. 2	Appendicitis	821	11.5	0(0)	Appendicitis	439	11.6	0(0)	Appendicitis	382	11.5	0(0)
No. 3	Diverticulitis & -osis	489	6.9	4(0.8)	Diverticulitis & -osis	293	7.7	4(1.4)	Diverticulitis & -osis	225	5.9	0(0)
No. 4	IBS & functional	482	11.5	8(1.7)	IBS & functional disorders	257	6.8	3(1.2)	IBS & functional disorders	196	6.8	5(2.2)
No. 5	Ileus	389	5.5	18(4.6)	Ileus	204	5.4	9(4.4)	Urolithiasis	192	5.8	0(0)
Chest pain												
No. 1	Ischemic heart diseases	1986	47.9	19(1.0)	Ischemic heart diseases	673	42.4	9(1.3)	Ischemic heart diseases	1313	51.4	10(0.8)
No. 2	Other heart diseases	922	22.3	28(3.0)	Other heart diseases	387	24.4	12(3.1)	Other heart diseases	535	20.9	16(3.0)
No. 3	Pneumonia	120	2.9	1(0.8)	Pulmonary embolism	57	3.6	0(0)	Pneumonia	80	3.1	1(1.3)

Frequencies and mortality of Top 5 discharge diagnoses (ICD-10 codes) specified by sex and presenting complaints in two out of three hospitals (continued)

	Total cohort	n	%	Mortality n(%)	Female	n	%	Mortality n(%)	Male	n	%	Mortality n(%)
No. 4	Pulmonary embolism	116	2.8	0(0)	Pneumonia	40	2.5	0(0)	Other respiratory diseases*	65	2.5	0(0)
No. 5	Other respiratory diseases*	103	2.5	0(0)	Other respiratory diseases*	38	2.4	0(0)	Pulmonary embolism	59	2.3	0(0)
Wounds												
No. 1	Injury of unspecified body region	780	19.5	1(0.1)	Injury of unspecified body region	246	17.1	1(0,1)	Injury of unspecified body region	534	20.8	0(0)
No. 2	Superficial wounds hand/fingers	359	9.0	0(0)	Superficial wounds head	123	8.6	0(0)	Superficial wounds hand/fingers	283	11.0	0(0)
No. 3	Superficial wounds head	333	8.3	0(0)	Infections of skin & subcutaneous tissue	92	6.4	0(0)	Superficial wounds head	210	8.2	0(0)
No. 4	Infections of skin & subcutaneous tissue	255	6.4	0(0)	Superficial wounds hand/fingers	76	5.3	0(0)	Infections of skin & subcutaneous tissue	163	6.4	0(0)
No. 5	Disorders of synovium and tendon	156	3.9	0(0)	Superficial wounds knee/lower leg	62	4.3	0(0)	Disorders of synovium and tendon	116	4.5	0(0)

Frequencies and mortality of Top 5 discharge diagnoses (ICD-10 codes) specified by sex and presenting complaints in two out of three hospitals (continued)

	Total cohort	n	%	Mortality n(%)	Female	n	%	Mortality n(%)	Male	n	%	Mortality n(%)
Trauma												
No. 1	Unspecified injuries of the head	698	24.3	4(0.6)	Unspecified injuries of the head	331	25.7	2(0.6)	Unspecified injuries of the head	367	23.2	2(0.5)
No. 2	Injury of unspecified body region	410	14.3	3(0.7)	Injury of unspecified body region	155	12.0	2(1.3)	Injury of unspecified body region	255	16.1	1(0.4)
No. 3	Injuries to the thorax	304	10.6	1(0.3)	Injuries to the thorax	111	8.6	0(0)	Injuries to the thorax	193	12.2	1(0.5)
No. 4	Superficial wounds head	220	7.7	3(1.4)	Superficial wounds head	98	7.6	2(2.0)	Superficial wounds head	122	7.7	1(0.8)
No. 5	Intracranial injury	197	6.9	5(2.5)	Intracranial injury	95	7.4	0(0)	Intracranial injury	102	6.5	5(4.9)
Collapse												
No. 1	Other heart disease	278	18.4	27(9.7)	Other heart disease	90	15.1	8(8.9)	Other heart disease	188	20.6	19(10.1)
No. 2	Unspecified injuries of the head	132	8.7	1(0.8)	Unspecified injuries of the head	58	9.7	0(0)	Unspecified injuries of the head	74	8.1	1(1.4)
No. 3	Epilepsy and recurrent seizures	87	5.8	0(0)	Epilepsy and recurrent seizures	30	5.0	0(0)	Epilepsy and recurrent seizures	57	6.2	0(0)
No. 4	Cerebral infarction	65	4.3	3(4.7)	Metabolic disorders	26	4.3	2(7.7)	Ischemic heart diseases	40	4.4	3(7.5)
No. 5	Diseases of inner ear	52	3.4	0(0)	Cerebral infarction	23	4.2	1(4.0)	Cerebral infarction	40	4.4	2(5.1)

Frequencies and mortality of Top 5 discharge diagnoses (ICD-10 codes) specified by sex and presenting complaints in two out of three hospitals (continued)

	Total cohort	n	%	Mortality n(%)	Female	n	%	Mortality n(%)	Male	n	%	Mortality n(%)
Palpitations												
No. 1	Other heart disease	2089	89.8	7(0.3)	Other heart disease	933	90.2	2(0.2)	Other heart disease	1156	98.5	5(0.4)
No. 2	Ischemic heart disease	60	2.6	0(0)	Ischemic heart disease	23	2.2	0(0)	Ischemic heart disease	37	2.9	0(0)
No. 3	Pneumonia	17	0.7	2(0.1)	Metabolic disorders	8	0.8	0(0)	Pneumonia	12	0.9	2(0.2)
No. 4	Metabolic disorders	14	0.6	0(0)	Hypertensive diseases	6	0.6	0(0)	Mental & behavioral disorders	8	0.6	1(12.5)
No. 5	Anemia	11	0.5	0(0)	Intestinal infectious diseases	5	0.5	0(0)	Other urinary diseases	7	0.5	1(14.3)
Headache												
No. 1	Cerebral infarction	162	15.2	4(2.5)	Cerebral infarction	78	13.9	4(5.1)	Cerebral infarction	84	16.9	0(0)
No. 2	Unspecified injuries of the head	112	10.6	0(0)	Migraine and other headache syndromes	63	11.2	0(0)	TIA/cerebrovascular disease	53	10.7	0(0)
No. 3	Migraine and other headache syndromes	104	9.8	0(0)	Unspecified injuries of the head	63	11.2	0(0)	Unspecified injuries of the head	49	9.9	0(0)
No. 4	TIA/cerebrovascular disease	88	8.3	0(0)	TIA/cerebrovascular disease	35	6.2	0(0)	Migraine and other headache syndromes	41	8.2	0(0)
No. 5	Intracranial hemorrhage	62	5.9	13(21.0)	Intracranial hemorrhage	34	6.0	9(26.5)	Intracranial hemorrhage	28	5.6	4(14.3)

Note. Discharge diagnoses are numbered 1 through 5 in order of appearance within both sex and presenting complaint. * Relates to diseases affecting the pleura or interstitium of the lung

SUPPLEMENTARY FILE 5

Patient characteristics of patients triaged as very urgent or most urgent.

	Male patients	Female patients
Demographics		
N (%)	20,838 (53.3)	18,222 (46.7)
Age, years, mean (SD)	60.7 (18.1)	60.8 (19.6)
Arrived by ambulance, N (%)	10,845 (59.2)	9290 (58.4)
Self-referral, N (%)	11,050 (53.7)	8614 (47.6)
Trauma/shock room, N (%)	2984 (21.5)	2429 (19.6)
Characteristics during ED-visit		
Top 10 presenting complaints, N (%)		
1. Feeling unwell	4173 (20.7)	3696 (20.7)
2. Chest pain	3676 (18.2)	2840 (15.9)
3. Dyspnoea	2924 (14.5)	2921 (16.4)
4. Abdominal pain	1574 (7.8)	1728 (9.7)
5. Extremity complaints	1123 (5.6)	1095 (6.1)
6. Collapse	1087 (5.4)	699 (3.9)
7. Trauma	1080 (5.3)	713 (4.0)
8. Palpitations	950 (4.7)	811 (4.6)
9. Intoxications	443 (2.2)	632 (3.5)
10. Headache	305 (1.5)	397 (2.2)
Miscellaneous	2843 (14.1)	2715 (15.2)
Vital score, N (%)		
No vital signs measured	25,188 (33.0)	22,221 (30.6)
≥1 vital signs measured	25,036 (32.8)	25,900 (35.7)
All vital signs measured	26,047 (34.2)	24,433 (33.7)
Vital signs		
SBP, mmHg, mean (SD)	130 (31)	132 (33)
DBP, mmHg, mean (SD)	85 (18)	83 (19)
Temperature, °C, median [IQR]	36.8 [36.4-37.3]	36.9 [36.5-37.4]
RR, /min, median [IQR]	18.0 [15-22]	15.0 [15-22]
SPO2, %, median [IQR]	98.0 [95-99]	98.0 [96-100]
HR, /min, mean (SD)	87 (25)	89 (24)

Patient characteristics of patients triaged as very urgent or most urgent.

	Male patients	Female patients
Demographics		
Pain scores, Numeric Rating Scale, N (%)		
Not measured	14,610 (70.1)	12,992 (71.3)
0	2237 (10.7)	1812 (9.9)
1-3	1546 (7.4)	1209 (6.6)
4-6	1252 (6.0)	1053 (5.8)
7-10	1193 (5.7)	1156 (6.3)
Biomarkers		
Sodium, median [IQR]	140 (138-142)	140 (137-142)
Creatinine, mmol/L, median [IQR]	87 [74-108]	68 [58-84]
Urea, mmol/L, median [IQR]	6.3 [4.9-8.5]	5.4 [4.1-7.4]
CRP, mg/L, median [IQR]	8 [3-41]	8 [3-34]
Leucocytes, x10^9/L, median [IQR]	9.2 [7.1-12.3]	9.2 [7.1-12.2]
Hb, mmol/L, mean (SD)	8.9 (1.4)	8.1 (1.1)
Lactate, mmol/L, median [IQR]	1.6 [1.1-2.5]	1.5 [1.0-2.2]

Abbreviations= N: number, SD: standard deviation, IQR: inter quartile range, ED: Emergency Department, SBP: systolic blood pressure, DBP: diastolic blood pressure, MAP: mean arterial pressure, RR: respiratory rate, HR: heart rate, SPO2: peripheral oxygen saturation, mmHg= millimetre mercury, °C: degrees Celsius, CRP: C-reactive Protein, Hb: hemoglobin.

SUPPLEMENTARY FILE 6

Sensitivity analysis 1

This sensitivity analysis assessed the association between sex and outcomes, using multivariable logistic regression models adjusted for age, presenting complaints, National Early Warning Score (NEWS), hospital location, radiology performed, fluid administration, number of consultations, urea, hemoglobin, and leukocytes.

	AOR (95%-CI)
In-hospital mortality	1.21 (1.11-1.31)
ICU/MCU admission	1.29 (1.22-1.38)
Total hospital admissions	1.15 (1.12-1.19)
ED-LOS (≥4h)	0.89 (0.86-0.91)
H-LOS (≥5d)	0.97 (0.93-1.00)
H-LOS (≥30d)	1.09 (0.96-1.25)

Abbreviations= AOR: adjusted odds ratio, CI: confidence interval, MCU: medium care unit, ICU: intensive care unit, ED-LOS: emergency department-Length Of Stay, H-LOS: hospital-Length Of Stay.
ORs >1 means a higher risk on the outcome for males compared to females.

Sensitivity analysis 2

In this sensitivity analysis patients who died in the ED were excluded, to assess whether sex differences in outcome were explained by patients in cardiac arrest in the ED.

A multivariable logistic regression model was used to assess the association between age and in-hospital mortality adjusted for age, top 10 chief complaints, hospital location, performed diagnostics (radiology, blood tests), fluid administration, number of consultations and vital signs.

	AOR (95%-CI)
In-hospital mortality	1.25 (1.15-1.36)

ORs >1 means a higher risk on the outcome for males compared to females.


EMERGENCY

6

THE ASSOCIATION BETWEEN SYSTOLIC BLOOD PRESSURE AND HEART RATE IN EMERGENCY DEPARTMENT PATIENTS: A MULTICENTRE COHORT STUDY

Bart GJ Candel
Wouter Raven
Søren Kabell Nissen
Marlies EB Morsink
Menno I Gaakeer
Mikkel Brabrand
Erik W van Zwet
Evert de Jonge
Bas de Groot

Published in the Journal of Emergency Medicine, 2023, Volume 65, Issue 1, p7-16

ABSTRACT

Background Guidelines and textbooks assert that tachycardia is an early and reliable sign of hypotension and an increased heart rate (HR) is believed to be an early warning sign for the development of shock, although this response may change by ageing, pain, and stress.

Objective To assess the unadjusted and adjusted associations between systolic blood pressure (SBP) and HR in Emergency Department (ED) patients of different age categories (18-50;50-80;>80 years).

Methods A multicentre cohort study using the Netherlands Emergency Department Evaluation Database (NEED) including all ED patients ≥18years from three hospitals in whom HR and SBP were registered at arrival to the ED. Findings were validated in a Danish cohort including ED patients. In addition, a separate cohort was used including ED patients with a suspected infection who were hospitalized from whom an SBP and HR were available before, during and after ED treatment. Associations between SBP and HR were visualized and quantified with scatterplots and regression coefficients (95%-CI).

Results A total of 81,750 ED patients were included from the NEED and a total of 2358 patients with a suspected infection. No associations were found between SBP and HR in any age category (18-50years: -0.03bpm/10mmHg, 95%CI -0.13-0.07, 51-80years: -0.43bpm/10mmHg, 95%CI -0.38- -0.50, >80years: -0.61bpm/10mmHg, 95%CI -0.53- -0.71)), nor in different subgroups of ED patient. No increase in HR existed with a decreasing SBP during ED treatment in ED patients with a suspected infection.

Conclusion No association between SBP and HR existed in ED patients of any age category, nor in ED patients who were hospitalized with a suspected infection, even during and after ED treatment. ED physicians may be misled by traditional concepts about HR disturbances because tachycardia may be absent in hypotension.

INTRODUCTION

Several textbooks and guidelines, such as the Advanced Trauma Life Support (ATLS) and Surviving Sepsis Campaign (SSC), assert that tachycardia is an essential part of the physiological response to hypotension.[39, 108-110] Consequently, tachycardia is often used as an early marker of hemodynamic instability.[39, 110] Nonetheless, in trauma and sepsis patients, hypotension is often not accompanied by tachycardia, and previous studies suggest not relying solely on heart rate (HR) disturbances to recognize a progressive fall in blood pressure.[111-116]

Even if heart rate increases with decreasing blood pressure, many factors, such as pain or anxiety, may cause a tachycardic response and thus heart rate may not be a reliable vital sign in an undifferentiated patient population such as in the Emergency Department (ED). More importantly, the heart rate response in situations with lowering blood pressure may be blunted with higher age, possibly by a decreased chronotropic beta-adrenergic responsiveness.[12, 117] For this reason, we hypothesized that an association between systolic blood pressure (SBP) and HR may exist in younger but is less clear or absent in older ED patients. A better understanding of the association between SBP and HR helps to assess the usefulness of HR as a vital sign to detect hypotension in acutely ill patients.

This study aims to assess the association between SBP and HR in the ED setting in different age categories.

METHODS

Study design and setting

This multicentre cohort study used three separate cohorts: the Netherlands Emergency department Evaluation Database (NEED), a quality registry for Dutch Emergency Departments (www.stichting-NEED.nl), the Danish Multicenter Cohort (DMC) to externally validate our findings, and a cohort including patients with a suspected infection (Sepsis op Spoedeisende Hulp; (SOS) cohort). The NEED contained data from three EDs, a tertiary care centre and two urban teaching hospitals (January 2017-September 2020). The DMC included all ED patients aged ≥18 years from two level II emergency departments with data between 2008 and 2013. Patients were consecutively sampled in relation to previous prospective studies at two level II emergency centers. This cohort was chosen for external validation to include a similar ED population as in the NEED, but in a different country and period. The cohort has been described in detail previously.[118, 119] The

SOS cohort included patients who were hospitalized from the ED with a suspected infection from two tertiary care centers and one urban teaching hospital in the Netherlands with data between 2011 and 2013. This was a retrospective analysis of prospectively collected data. The cohort has been described in detail previously.[120] This cohort enabled the exploration of vital signs within one patient during resuscitation, as it contained vital sign registrations at three different time points in the ED.

Patients were stratified by age categories (18-50, 51-80, >80 years). These age categories were chosen because the incidence of hypertension is low till the age of 50 years in the Netherlands after which it gradually increases (www.cbs.nl). The cut-off at 80 years has been used in previous studies.[82, 119]

Participants
In the NEED and DMC, all consecutive patients ≥18 years presenting to the ED were included if both SBP and HR were registered.

The SOS cohort included consecutive ED patients aged 17 years and older, with suspected infection and an urgent to very urgent triage category according to the Manchester triage system, who were admitted to the hospital and treated with intravenous antibiotics. For the current study, patients of 17years old were excluded from the database.

Data collection
Previously, we described in detail how vital signs were measured and which data were collected in the NEED.[82] In summary, we collected demographic variables, disease severity, proxies for comorbidity and complexity, vital signs, laboratory results and outcomes. One set of vital signs (HR, SBP, peripheral oxygen saturation, respiratory rate, temperature, diastolic blood pressure) was recorded per patient measured at the beginning of ED presentation before ED treatment. Vital signs were not recorded if this was not necessary for the ED presentation. This was similar for the DMC. In the SOS, vital signs were registered before ED treatment (Time 0h at arrival identical to the NEED cohort), during ED treatment (approximately around time 30 minutes) and after ED treatment (before admission to the ward). During ED stay, the lowest SBP and the highest HR value were registered. Also, the use of beta-blocking agents and calcium antagonists was registered only in this cohort.

Data analyses

Descriptive analyses

Data were presented as mean (standard deviation (SD)) if normally distributed and median (interquartile range (IQR)) if skewed.

Main statistical analyses

Univariable associations between SBP and HR were investigated and visualized using scatterplots for three age categories. Both a linear and locally estimated scatterplot smoothing (LOESS) fit with 95% confidence intervals (95%CI) were fitted to the data. In the figures, the size of the dots indicates the precision of the estimate for observed heart rate and is based on the inverse of the standard deviation. The larger, the higher the precision. A linear regression coefficient with 95%CI and a R squared were also presented.

To correct the association between SBP and HR for other vital sign values and pain scores, an additional analysis was performed. We investigated the association between SBP and HR using the R package mgcv, to fit a generalized additive regression model (GAM) adjusted for other vital signs and pain scores which were left constant at 'normal' values (respiratory rate (18/min.), peripheral oxygen saturation (97%), temperature (36 degrees Celsius), numeric rating scale for pain (NRS) (0)), on the NEED data. Missing NRS was considered as a score of 0. If other vital signs were missing the patient was excluded. This analysis was intended to be descriptive, and therefore we did not attempt formal statistical inference in terms of p-values and confidence intervals.

In the SOS we assessed whether the use of beta-blocking agents or calcium antagonists affected our results by linear fits for patients with beta-blocking agents and patients without beta-blocking agents. A linear regression coefficient with a p-value and an R squared was presented. In addition, we assessed whether deterioration of blood pressure was accompanied by an increasing HR. We therefore calculated the deltaSBP = SBP after ED treatment – SBP before ED treatment, and similarly a deltaHR. We assessed the relationships between those two variables using a scatterplot with linear fit.

Subgroup analyses

We performed four subgroup analyses in the NEED. The first subgroup analysis included all patients with an SBP<120mmHg, to put more focus on the hypotensive patients in this study. In this subgroup, we compared the

in-hospital mortality rates in patients with a low HR response (below the median) to hypotension and with a high HR response (above the median). In this way we could study whether the absence of a tachycardic response had a poor outcome compared to patients with a tachycardic response.

To assess whether the aetiology of a low SBP and severity of illness affected the univariable association with HR, the relationship between SBP and HR were also investigated for the subgroups suspected infection and trauma patients (potentially having haemorrhage) in the NEED, and for patients admitted to the Intensive Care Unit (ICU). In the subgroups suspected infection and trauma, the most likely cause for hypotension is septic shock and hemorrhagic shock, respectively, while other reasons are less likely. In this way, we could assess the association between SBP and HR in two types of shock patients. Patients with a suspected infection were selected in the NEED if blood cultures were obtained and if they were hospitalized. Trauma patients were selected if they were triaged as very urgent or immediate (according to the Manchester Triage System or Dutch Triage Standard) and by the following triage presenting complaints: Severe trauma; facial complaints; fall; extremity complaints; large scale incidents; neck pain; thorax injury; head injury; wounds. Additionally, a radiological test had to be performed. Patients admitted to the ICU were registered in the NEED and were studied as last subgroup.

All statistical analyses were performed in Rstudio (packages rms, dplyr, and mgcv).

RESULTS

Patient characteristics

The NEED contained 148,828 patients ≥18 years. In total, 81,750 (54.9%) patients could be included in whom both SBP and HR were registered. See a flow diagram in figure 1. The mean age of patients in the NEED was 61 years (SD 19). Mean SBP was higher in older patients >80 years with 141mmHg (SD 34) compared to the youngest age category with mean SBP of 124mmHg (SD 26). Patient characteristics are described in table 1.

Table 1: Patient characteristics from the NEED

NEED cohort	18-50years (N=22617)	51-80years (N=46350)	>80years (N=12783)	All (N=81750)
Age, years				
Mean (SD)	35 (9.8)	6 (8.4)	86 (3.9)	61 (19)
Sex				
female	12170 (53.8%)	21054 (45.4%)	7144 (55.9%)	40368 (49.4%)
Triage category, N (%)				
immediate	1044 (4.6%)	3050 (6.6%)	788 (6.2%)	4882 (6.0%)
very urgent	6534 (28.9%)	15037 (32.4%)	3891 (30.4%)	25462 (31.1%)
urgent	9771 (43.2%)	19349 (41.7%)	5362 (41.9%)	34482 (42.2%)
non-urgent	4885 (21.6%)	8181 (17.7%)	2516 (19.7%)	15582 (19.1%)
Missing	383 (1.7%)	733 (1.6%)	226 (1.8%)	1342 (1.6%)
Systolic Blood Pressure (mmHg)				
Mean (SD)	124 (26)	135 (32)	141 (34)	133 (31)
Heart rate (bpm)				
Mean (SD)	87 (20)	86 (22)	83 (21)	86 (21)
Respiratory Rate (/min.)				
Mean (SD)	17 (5)	18 (5)	19 (6)	18 (5)
Missing	6865 (30.4%)	9729 (21.0%)	2385 (18.7%)	18979 (23.2%)
Peripheral oxygen saturation (%)				
Mean (SD)	98.4 (2.5)	96.7 (3.6)	96.0 (3.8)	97.1 (3.5)
Missing	823 (3.6%)	1819 (3.9%)	509 (4.0%)	3151 (3.9%)
Temperature (degrees Celcius)				

Table 1: Patient characteristics from the NEED (continued)

NEED cohort	18-50years (N=22617)	51-80years (N=46350)	>80years (N=12783)	All (N=81750)
Mean (SD)	37.1 (0.8)	37.0 (0.9)	36.9 (0.9)	37.0 (0.9)
Missing	4573 (20.2%)	7944 (17.1%)	2312 (18.1%)	14829 (18.1%)
Numeric Rating scale for pain (1 to 10)				
Median (IQR)	4 (1-6)	2 (0-5)	2 (0-4)	3 (0-5)
Missing	13088 (57.9%)	30002 (64.7%)	8698 (68.0%)	51788 (63.3%)
Supplemental Oxygen				
yes	429 (1.9%)	2879 (6.2%)	1353 (10.6%)	4661 (5.7%)
Missing	6086 (26.9%)	10745 (23.2%)	2237 (17.5%)	19068 (23.3%)
Fluid administration (ml)				
0cc	11407 (50.4%)	26027 (56.2%)	7289 (57.0%)	44723 (54.7%)
0-500cc	2031 (9.0%)	4314 (9.3%)	1322 (10.3%)	7667 (9.4%)
>500cc	2320 (10.3%)	5159 (11.1%)	1239 (9.7%)	8718 (10.7%)
Missing	6859 (30.3%)	10850 (23.4%)	2933 (22.9%)	20642 (25.3%)
In-hospital mortality, N(%)				
died	90 (0.4%)	1197 (2.6%)	770 (6.0%)	2057 (2.5%)
Missing	188 (0.8%)	688 (1.5%)	264 (2.1%)	1140 (1.4%)
ICU admission, N (%)				
ICU admission	350 (1.5%)	924 (2.0%)	140 (1.1%)	1414 (1.7%)
Missing	318 (1.4%)	397 (0.9%)	66 (0.5%)	781 (1.0%)

Abbreviations= N: number, SD: standard deviation, IQR: interquartile range, GCS: Glasgow Coma Scale, ICU: Intensive Care Unit

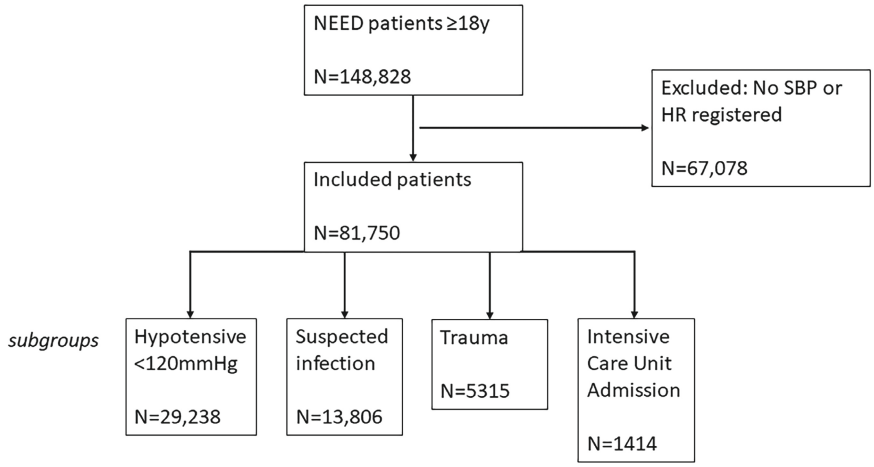


Figure 1 Flow diagram throughout the NEED cohort.

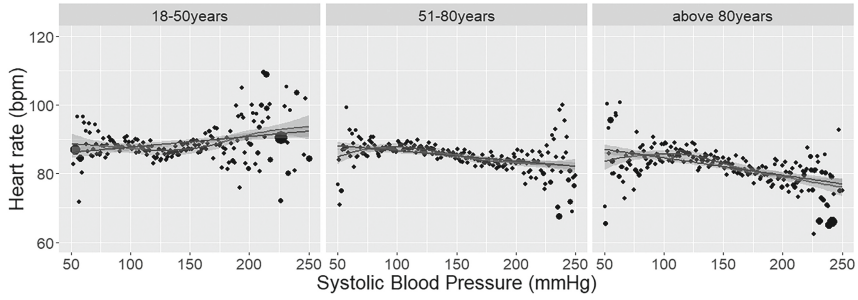


Figure 2 The univariable relationship between systolic blood pressure (SBP) and heart rate (HR) stratified by age categories in the NEED cohort (N=81,750). The size of the dots indicates the precision of the estimate for observed heart rate and is based on the inverse of the standard deviation. The larger, the higher the precision. Both a linear and locally estimated scatterplot smoothing (LOESS) fit with 95% confidence intervals were fitted to the data. No association was found in any age category between SBP and HR.

Main results

Figure 2 demonstrates the association between SBP and HR in three age categories at arrival in the ED. No association between HR and SBP was found in any age group. No association between SBP and HR was found in any subgroup of hypotensive patients, trauma patients, patients with a suspected infection, and ED patients admitted to the ICU (see figure 3). Adjustment for other vital signs and pain scores did not affect the association between SBP and HR (Figure 4).

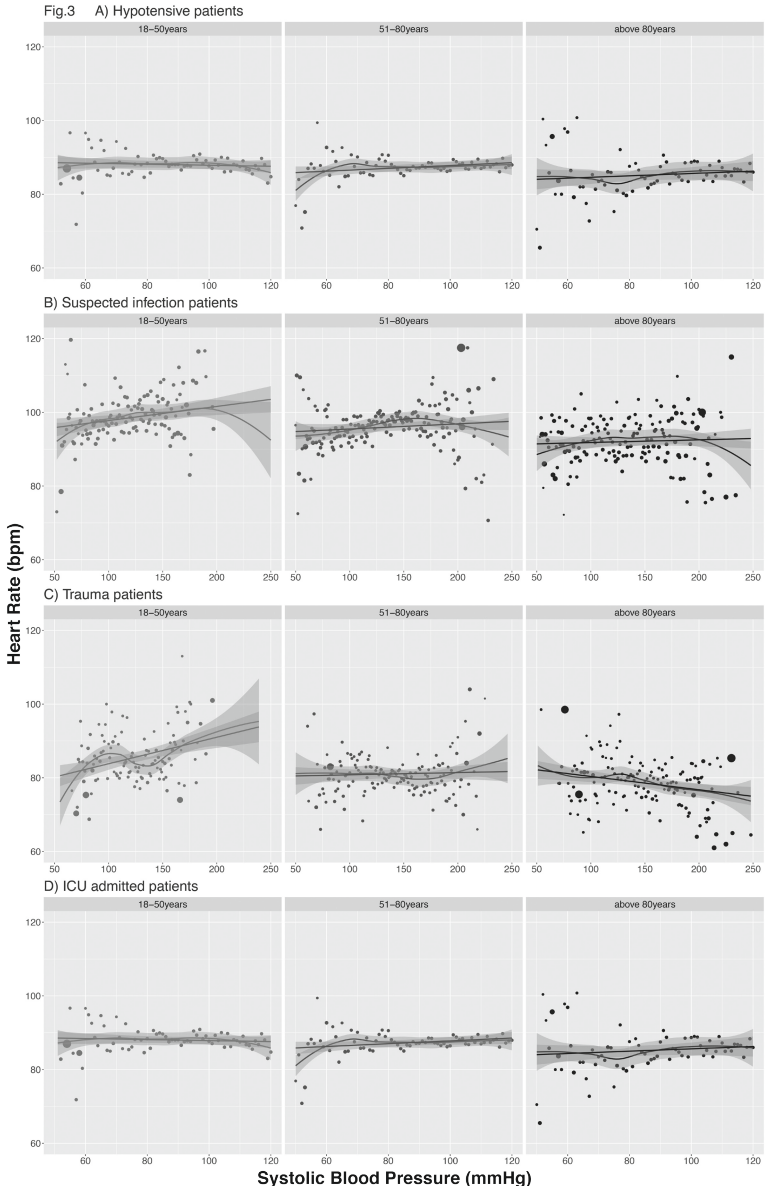


Figure 3. Subgroup analyses in the NEED registry. Scatterplots are presented for the association between systolic blood pressure and heart rate in different age categories. Panel A shows the association between systolic blood pressure (SBP) and heart rate in a subgroup of patients with an SBP <120mmHg (N=29,238). Panel B shows a subgroup of patients with a suspected infection (N=13,806). Panel C shows a subgroup of trauma patients (N=5315). Panel D shows a subgroup of ED patients who were admitted to the Intensive Care Unit (N=1414). No association was found in any age category between SBP and HR.

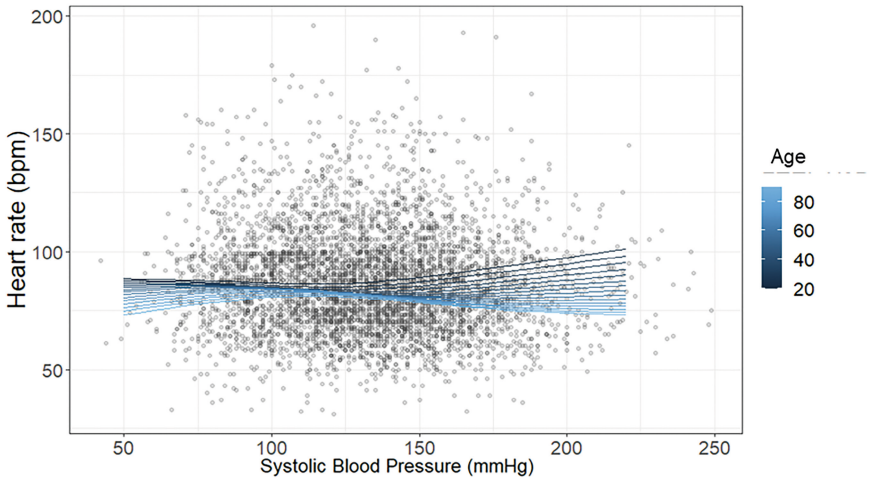


Figure 4 The associations between systolic blood pressure, heart rate, and age are shown using a generalized additive logistic regression model, adjusted for temperature, peripheral oxygen saturation, pain score and respiratory rate.

The SOS cohort included 2358 patients with a suspected infection. Patient characteristics are described in supplementary file 1. No clear increase in HR could be found with decreasing SBP for any age category neither before, during or after ED treatment (Figure 5). The use of beta-blocking agents or calcium antagonists did not affect the relationship. If the vital signs after ED treatment of an individual patient were compared with the vital signs before ED treatment, no increase in deltaHR existed with a decreasing SBP during resuscitation in the ED (e.g., a negative deltaSBP; see supplementary file 2).

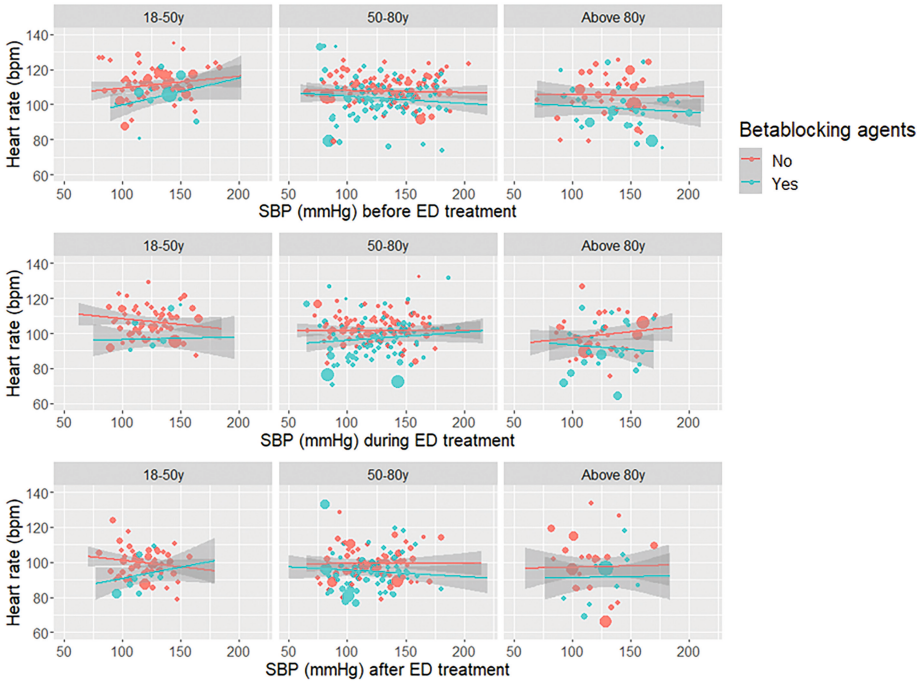


Figure 5 The univariable relationship between SBP and HR stratified by age categories and time of registration in a cohort of patients with a suspected infection (N=2358). A linear line was fitted to the data. The relationship was assessed before ED treatment (T0), during ED treatment (T30min.) and after ED treatment just before ward admission. Patients with beta-blocking agents or calcium antagonists were marked. No association was found in any age category between SBP and HR.

External validation in a Danish cohort of ED patients showed similar results (supplementary file 3). No clear increase in HR could be found with decreasing SBP for any age category.

Regression coefficients and R squared for the different cohorts and subgroups are presented in supplementary file 4. Although some regression coefficients were statistically significant, they were close to zero and therefore not clinically relevant.

In the subgroup with hypotensive patients (SBP <120mmHg) we compared mortality rates of patients with a low HR and patients with a high HR response. Mortality was higher in the patients with a high HR (see

supplementary file 5). However, the hypotensive patients >80years still had a 7.3% mortality risk in the absence of a tachycardic response.

DISCUSSION

In this multicentre cohort study, no association between SBP and HR was found in ED patients of any age category or subgroup including suspected infection, trauma and ICU admitted patients. In the ED, tachycardia alone may therefore not be helpful in detecting hypotension.

Although several textbooks and guidelines assert that tachycardia is an early and reliable sign of hypotension to maintain sufficient cardiac output and circulating volume,[39, 108-110, 121] low blood pressure was not accompanied by an increase in HR in the present study. Textbooks are based on theoretic physiological principles explaining that compensated hypotension is characterized by intense sympathetic stimulation of the circulation elicited by baroreceptor reflexes.[108, 109, 121, 122] However, several animal studies have shown an absence of tachycardia in haemorrhage, which is in line with the findings of the present study.[122-125] The absence of tachycardia as a response to hypotension during haemorrhage has been explained to be a vagal reflex.[114, 125-127]

In sepsis it is currently believed that inflammation leads to vasodilation which alters cardiac output due to preload reduction.[39, 128, 129] As a result, sympathetic activation is triggered and should lead to a compensated tachycardia and vasoconstriction. Nevertheless, decreasing blood pressure was not accompanied by an increased HR in younger and older patients on average cohort level. This finding follows several studies which have demonstrated that an altered heart rate variability in sepsis is common and correlates with the severity of the systemic infection with a higher mortality rate.[115, 116]

At the start of this study we hypothesized that an association between SBP and HR could be absent in older patients due to the decreasing chronotropic response with increasing age.[12, 117] However, the association was also lacking in younger patients. Our results were confirmed in a Danish multicentre cohort, implying that they can be generalized to other ED settings and times. Also, we have shown that adjusting for other vital signs and pain did not affect the association between SBP and HR in ED patients. We can only speculate why an increased HR did not accompany hypotension in the ED. A potential explanation is the large variety in ED presentations. While patients with deranged hemodynamics, such as septic patients, would be

expected to have a combination of tachycardia and low blood pressure, other ED patients presenting with pain or anxiety disorders could have tachycardia with high blood pressure. Other factors such as changes in blood volume, pain, anxiety, parasympathetic and sympathetic stimuli and circulating hormones are likely explanations that may play an essential role in ED patients. Because of this heterogeneity, no association between SBP and HR may have been found.

The NEED only contained vital sign registrations at arrival to the ED, but repeated analyses in a cohort with multiple registrations during and after ED treatment showed similar results. Remarkably, a decrease in SBP during and after resuscitation was not accompanied by an increase in HR in individual patients with a suspected infection.

The importance of our observations lies in the fact that hypotension cannot be recognized solely by the presence of tachycardia, and so HR may be an unreliable vital sign to screen for physiological derangement and hypotension in ED patients of all age categories, while guidelines such as the ATLS continue to use HR as one of the vitals to classify the degree of haemorrhagic shock.[39, 108-110] In this study we demonstrated that mortality rates are lower in hypotensive patients without a tachycardic response than in patients with a tachycardia. Hypothetically, the patients with an SBP below 120mmhg and absent tachycardic response might not have essential hypotension, e.g., a perfusion problem. Nonetheless, the in-hospital mortality rate was still substantially high in older patients, possibly caused by a decreased chronotropic beta-adrenergic responsiveness or by beta-blocking medication. Therefore ED physicians should not be misled by traditional concepts about HR disturbances.[39, 108-110]

Despite several strengths like the large sample size and the fact that we performed analyses in multiple independent cohorts, several limitations merit emphasis. First, only one set of vital signs at ED arrival was available in the NEED database, which may have affected our results. Possibly the first set of vital signs is affected by measurement artifacts, or many more confounders, such as mental stress and pain affecting HR, which could blur the association with SBP. However, we assessed the relationship between SBP and HR in a cohort including patients with a suspected infection at three different times during ED stay (supplementary file 2), which showed similar results. Secondly, we could not adjust the analyses for potential confounders such as anxiety or vasopressor/inotropic agents. Nonetheless, we have shown no clear relationship between SBP and HR even after adjusting for vital signs and pain. Finally, we assessed the relationship between SBP

and HR at cohort level. A decreasing SBP may still be accompanied by an increasing HR in individual patients, even though no increased HR was found with a decreasing SBP in patients with suspected infection in the SOS cohort.

In summary, the present study indicates that an association between SBP and HR cannot be detected in ED patients of any age category or subgroup. Acute care guidelines should acknowledge that HR may not be a reliable vital sign to detect physiological derangement in the ED.

SUPPLEMENTARY FILE 1

Patient characteristics from the SOS multicenter cohort including patients with suspected infection

Patients with suspected infection	18-50y (N=573)	50-80y (N=1497)	Above 80y (N=288)	All (N=2358)
Age, years, median (IQR)	40 (31-46)	66 (60-72)	85 (82-88)	64 (51-74)
Sex, female	503 (49.4%)	593 (39.4%)	120 (41.2%)	1001 (42.5%)
Systolic Blood Pressure (mmHg) before ED treatment				
Median [IQR]	125[113-145]	132 [115-151]	135 [110-155]	130 [114-149]
Missing	94 (9.2%)	140 (9.3%)	10 (3.4%)	244 (8.7%)
Systolic Blood Pressure (mmHg) during ED treatment				
Median [IQR]	114 [72.5-130]	123 [106-140]	122 [103-142]	121 [102-137]
Missing	511 (50.2%)	418 (27.8%)	87 (29.9%)	1016 (36.1%)
Systolic Blood Pressure (mmHg) after ED treatment				
Median [IQR]	118 [107-130]	120 [106-135]	120 [107-138]	120 [106-134]
Missing	723 (71.0%)	701 (46.5%)	136 (46.7%)	1560 (55.4%)
Heart rate (bpm) before ED treatment				
Median [IQR]	93.0 [3.00-115]	107 [94.0-120]	102 [90.0-115]	104 [84.0-117]
Missing	15 (1.5%)	29 (1.9%)	2 (0.7%)	46 (1.6%)
Heart rate (bpm) during ED treatment				
Median [IQR]	98.0 [70.0-111]	100 [87.0-112]	96.0 [82.0-110]	99.0 [85.0-111]
Missing	495 (48.6%)	389 (25.8%)	79 (27.1%)	963 (34.2%)

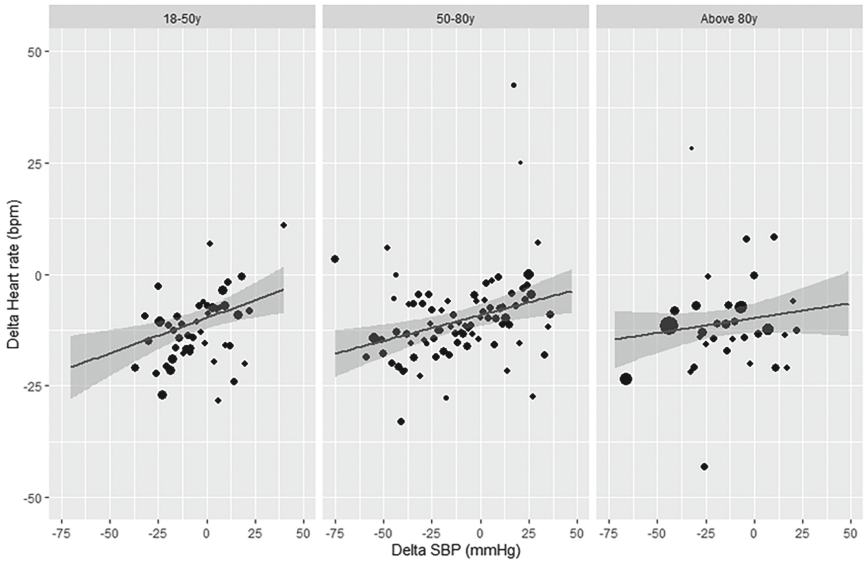
Patient characteristics from the SOS multicenter cohort including patients with suspected infection (continued)

Patients with suspected infection	18-50y (N=573)	50-80y (N=1497)	Above 80y (N=288)	All (N=2358)
Heart rate (bpm) after ED treatment				
Median [IQR]	99.0 [90.0-108]	96.0 [85.0-109]	96.0 [80.0-108]	97.0 [85.0-108]
Missing	701 (68.9%)	627 (41.6%)	130 (44.7%)	1458 (51.8%)
Use of betablocking agents or calcium antagonists				
Yes	69 (12.0%)	540 (36.1%)	125 (43.0%)	731 (31%)
Missing	2 (0.3%)	11 (0.7%)	3 (1.0%)	16 (0.7%)
Use of other Antihypertensiva				
Yes	73 (12.7%)	568 (37.9%)	150 (52.1%)	791 (30.0%)
Missing	1 (0.2%)	11 (0.7%)	3 (1.0%)	15 (0.6%)
Fluid administration in the ED				
0ml	60 (10.5%)	166 (11.1%)	32 (11.1%)	258 (10.9%)
0-1000ml	159 (27.7%)	410 (27.4%)	100 (34.7%)	669 (28.4%)
1000-2000ml	282 (49.2%)	682 (45.6%)	115 (39.9%)	1079 (45.8%)
>2000ml	72 (12.6%)	239 (16.0%)	41 (14.2%)	352 (14.9%)
In-hospital mortality				
Yes	13 (2.3%)	96 (6.4%)	39 (13.5%)	148 (6.3%)
Missing	4 (0.7%)	11 (0.7%)	6 (2.1%)	21 (0.9%)

Abbreviations= N: number, SD: standard deviation, IQR: interquartile range, GCS: Glasgow Coma Scale, ICU: Intensive Care Unit, mmHg: Millimeter Mercury.

SUPPLEMENTARY FILE 2

The association between delta SBP and delta HR



The association between delta systolic blood pressure and delta heart rate in patients with suspected infection. Delta SBP was calculated as the SBP after ED treatment – SBP before ED treatment. Delta heart rate was calculated as the HR after ED treatment – HR before ED treatment.
If the SBP is lower after ED treatment (a negative delta SBP), the heart rate does not respond with a positive delta HR in individual patients as can be seen in the figure.

SUPPLEMENTARY FILE 3

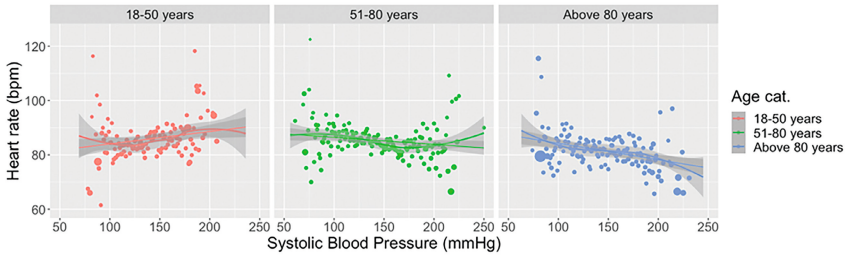
External validation in Denmark: To validate our findings in a different country and period, a Danish multicentre Cohort (DMC) was used, including patients from two level II emergency departments with data between 2008 and 2013.

Patient characteristics

DMC	18-50 (N=4724)	51-80 (N=7456)	>80 (N=2462)	All (N=14642)
Age				
Mean (SD)	35.5 (9.78)	66.8 (8.32)	86.1 (4.09)	60.0 (20.0)
Gender				
female	2096 (44.4%)	3820 (51.2%)	1131 (45.9%)	7047 (48.1%)
Systolic blood pressure (mmHg)				
Mean (SD)	132 (18.9)	140 (25.4)	140 (27.3)	137 (24.1)
Pulse (/min)				
Mean (SD)	84.4 (19.1)	84.9 (20.3)	81.6 (19.6)	84.2 (19.8)
inhospital mortality				
Yes	11 (0.2%)	185 (2.5%)	159 (6.5%)	355 (2.4%)
Admission to Intensive Care Unit				
Yes	100 (2.1%)	233 (3.1%)	72 (2.9%)	405 (2.8%)

Abbreviations= N: number, SD: standard deviation, DMC: Danish Multicenter Cohort, mmHg: Millimeter Mercury.

The figure below shows the association between SBP and HR for the external validation cohort.



SUPPLEMENTARY FILE 4

Correlation coefficients: For each cohort or subgroup, the correlation coefficient is presented as a change in heart rate (bpm) per 10mmHg increase in Systolic blood pressure is presented for the linear fit on the data for all patients with an SBP between 50 and 250mmHg. Also, the R squared is presented for the linear fit.

	Change in HR (bpm) per 10mmHg increase in SBP (95%-CI)	R squared
NEED cohort (all patients, N=81,750)		
18-50y	-0.031 (-0.13- 0.07)	0,000
51-80y	-0.43 (-0.38- -0.50)	0.004
>80y	-0.61 (-0.53- -0.71)*	0.010
NEED cohort (subgroup trauma patients, N=5315)		
18-50y	0.34 (-0.05- 0.72)	0.003
51-80y	-0.10 (-0.31- 0.10)	0.000
>80y	-0.38 (-0.65- -0.12)*	0.007
NEED cohort (subgroup suspected infection, N= 13806)		
18-50y	0.54 (0.26-0.82)*	0.005
51-80y	0.43 (0.29-0.58)*	0.004
>80y	0.15 (-0.11-0.40)	0.001
NEED cohort (ICU admitted patients, N= 1414)		
18-50y	-0.59 (-1.4-0.23)	0.007
51-80y	0.30 (-0.13-0.74)	0.001
>80y	-0.73 (-1.7-0.28)	0.015
DMC cohort (All patients, N=14642)		
18-50y	0.64 (0.35-0.92)*	0.004
51-80y	-0.55 (-0.73- -0.37)*	0.005
>80y	-0.53 (-0.80- -0.26)*	0.005

Abbreviations= N: number, 95%-CI: 95% Confidence Intervals, ICU: Intensive Care Unit, NEED: Netherlands Emergency department Evaluation Database, DMC: Danish Multicenter Cohort, bpm: beats per minute, SBP: systolic blood pressure, HR: heart rate.
*Statistically significant correlation coefficient ($p<0.05$)

SUPPLEMENTAL FILE 5.

Patients with systolic blood pressure below 120 mmHg stratified by heart rate below or above median.

	18-50y		51-80y		>80y		Total	
Number	**LOW HR (N=5056)**	**HIGH HR (N=4787)**	**LOW HR (N=8164)**	**HIGH HR (N=7614)**	**LOW HR (N=1879)**	**HIGH HR (N=1738)**	**LOW HR (N=15099)**	**HIGH HR (N=14139)**
Age								
Median [IQR]	35 [27-44]	33 [25-43]	67 [59-73]	66 [59-73]	85 [83-88]	85 [83-88]	61 [44-74]	61 [43-73]
Sex								
male	2088 (41.3%)	2004 (41.9%)	4754 (58.2%)	4195 (55.1%)	986 (52.5%)	875 (50.3%)	7828 (51.8%)	7074 (50.0%)
Heart rate (bpm)								
Median [IQR]	74 [66-80]	98 [92-110]	73 [65-80]	99 [93-110]	72 [64-78]	97 [91-110]	73 [65-80]	98 [92-110]
Systolic blood pressure (mmHg)								
Median [IQR]	105 [92-114]	102 [91-112]	103 [90-112]	102 [90-112]	103 [90-113]	104 [91-113]	103 [91-113]	103 [91-112]

Supplemental file 5. Patients with systolic blood pressure below 120 mmHg stratified by heart rate below or above median. (continued)

	18-50y		51-80y		>80y		Total	
Number	**LOW HR (N=5056)**	**HIGH HR (N=4787)**	**LOW HR (N=8164)**	**HIGH HR (N=7614)**	**LOW HR (N=1879)**	**HIGH HR (N=1738)**	**LOW HR (N=15099)**	**HIGH HR (N=14139)**
In-hospital mortality								
died	19 (0.4%)	38 (0.8%)	207 (2.5%)	394 (5.2%)	138 (7.3%)	184 (10.6%)	364 (2.4%)	616 (4.4%)
Missing	34 (0.7%)	68 (1.4%)	233 (2.9%)	242 (3.2%)	98 (5.2%)	76 (4.4%)	365 (2.4%)	386 (2.7%)
Intensive Care Unit Admission								
ICU admission	49 (1.0%)	145 (3.0%)	192 (2.4%)	295 (3.9%)	29 (1.5%)	44 (2.5%)	270 (1.8%)	484 (3.4%)
Missing	76 (1.5%)	65 (1.4%)	82 (1.0%)	57 (0.7%)	13 (0.7%)	7 (0.4%)	171 (1.1%)	129 (0.9%)

NEED: Netherlands Emergency department Evaluation Database, HR: Heart rate, bpm: beats per minute, mmHg: millimeter Mercury, ICU: Intensive Care Unit, IQR: Interquartile Range.
The table includes patients with a Systolic Blood Pressure <120mmHg (N=29,238). Patients were divided in a low heart rate response below the median heart rate and a high heart rate response above the median heart rate. Median heart rate was 86bpm for both the age-group 18-50years and 51-80years, and 84bpm for the >80years.



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.

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 $R^2_{Cox\text{-}Schnell}$ (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

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).

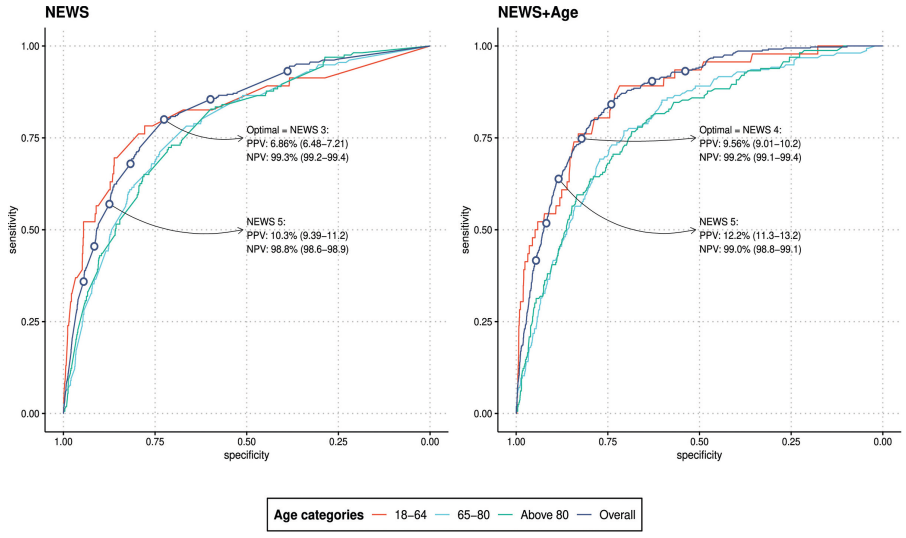


Figure 1 Receiver operating characteristics for NEWS and NEWS+Age by age groups and overall in the Danish Multicenter Cohort for prediction of in-hospital mortality. Values in parentheses indicate the 95% confidence intervals.

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

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).

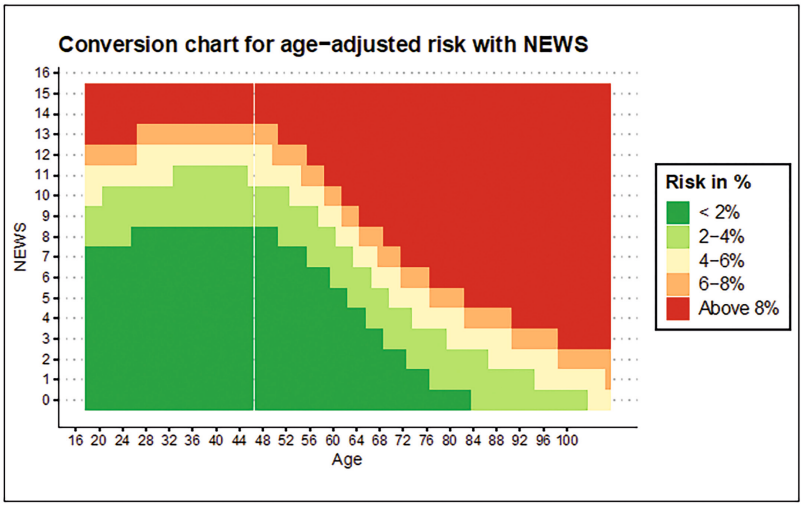


Figure 2 Conversion-chart for calculation of risk of in-hospital mortality with NEWS and age for for risk-categories <2%, 2-4%, 4-6%, 6-8% and >8%. Using this conversion chart, a 76-year-old carries a risk below 6% and an 84-year-old carries a risk of 6-8% with the same NEWS of 4.

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).

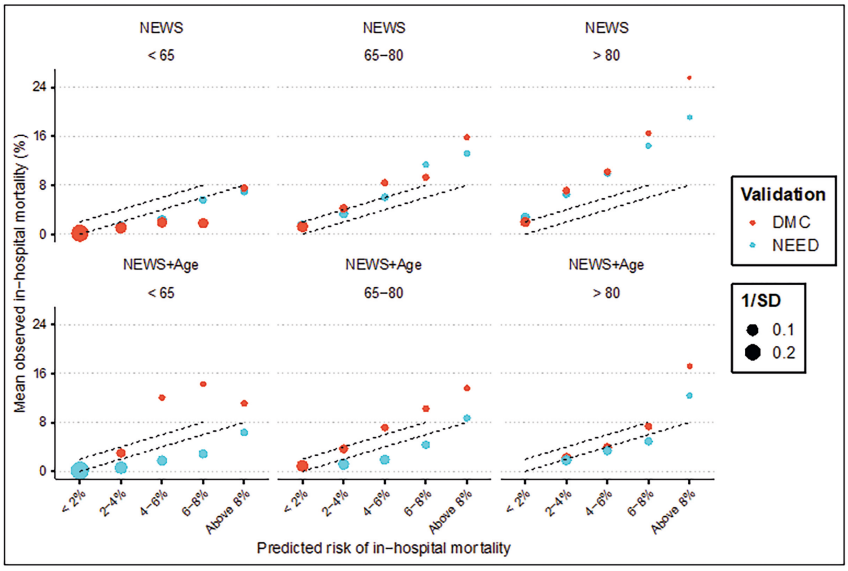


Figure 3 Internal and external calibration of the conversion-chart. The predicted risk of in-hospital mortality has been categorized to categories <2%, 2-4%, 4-6% and 6-8 in the conversion chart. Dots placed within dotted lines represent ideal calibration. The size of the dots indicate the precision 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

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

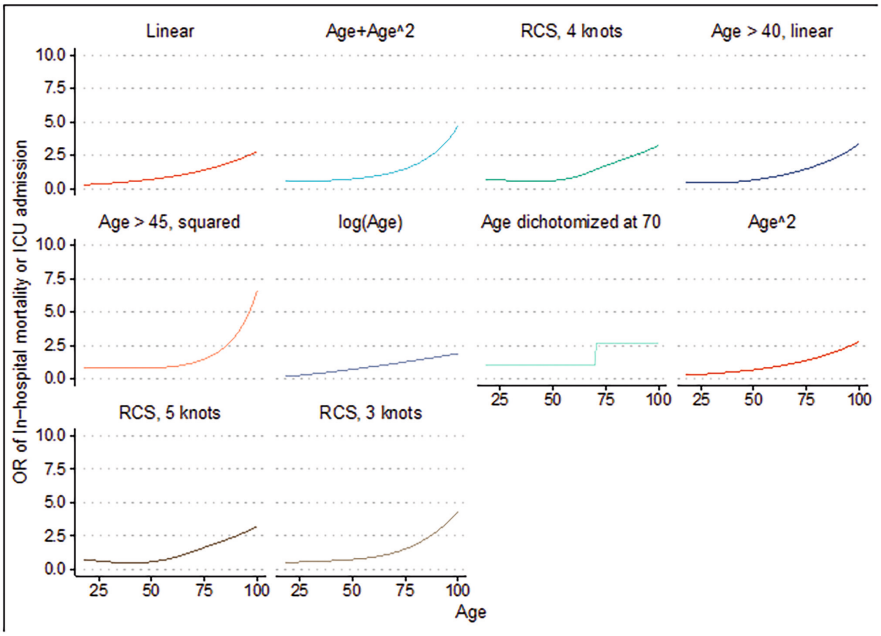
Univariate regression fits of age to ICU admission or in-hospital mortality according to transformation-term and derived measures of fit

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	

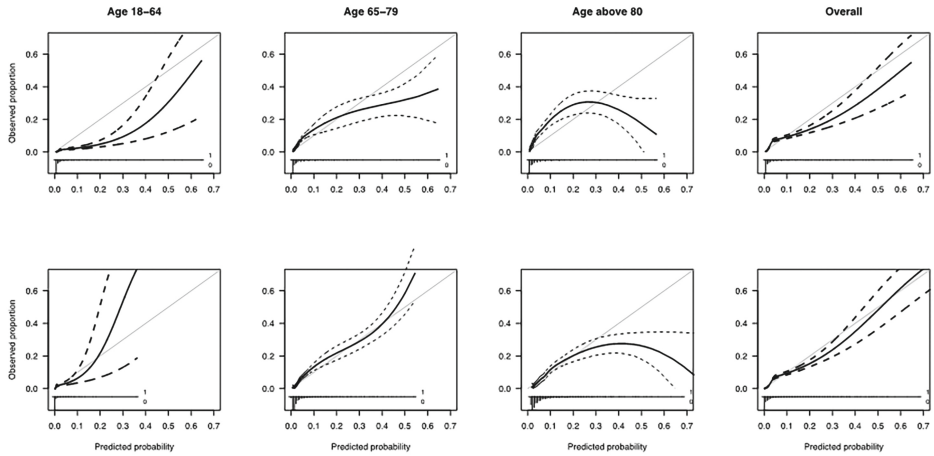
Abbreviations: S.E., Standard Error; NEWS, National Early Warning Score; RCS, Restricted Cubic Splines; Pr, Probability; AUROC, Area Under Receiver Operating Characteristic; CI, Confidence Interval.

[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

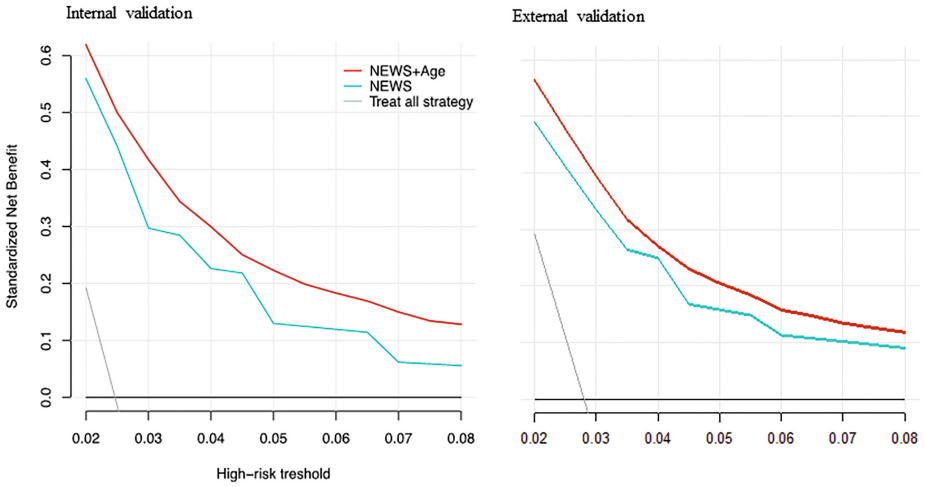
Flexible calibration plots for NEWS and NEWS+Age by age groups and overall, in the DMC. Bold lines are flexible calibration fits. A non-parametric loess smoother is used instead of restricted cubic in age categories 65-80 and above 80 due to number of observations being below 5000. Grey lines represent 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


EMERGENCY

8

DEVELOPMENT AND EXTERNAL VALIDATION OF THE INTERNATIONAL EARLY WARNING SCORE FOR IMPROVED AGE AND SEX ADJUSTED IN-HOSPITAL MORTALITY PREDICTION IN THE EMERGENCY DEPARTMENT.

Bart GJ Candel
Søren Kabell Nissen
Christian H. Nickel
Wouter Raven
Wendy Thijssen
Menno I Gaakeer
Annmarie Touborg Lassen
Mikkel Brabrand
Ewout W Steyerberg
Evert de Jonge
Bas de Groot

Published in Critical Care Medicine – 2023, Volume 51, Issue 7, p881–891

ABSTRACT

Objective Early Warning Scores (EWS) have a great potential to assist clinical decision making in the Emergency Department (ED). However, many EWS contain methodological weaknesses in development and validation and have poor predictive performance in older patients. The aim of this study was to develop and externally validate an International Early Warning Score (IEWS) based on a recalibrated NEWS model including age and sex and evaluate its performance independently at arrival to the ED in three age categories (18-65y; 66-80y; >80years).

Design and setting International multicenter cohort study using data from three Dutch EDs. External validation was performed in two EDs in Denmark.

Patients All consecutive ED patients ≥18years in the Netherlands Emergency department Evaluation Database (NEED) with at least two registered vital signs were included, resulting in 95,553 patients. For external validation, 14,809 patients were included from a Danish Multicenter Cohort (DMC).

Measurements Model performance to predict in-hospital mortality was evaluated by discrimination, calibration curves and summary statistics, reclassification, and clinical usefulness by decision curve analysis.

Main results In-hospital mortality rate was 2.4% (N=2314) in the NEED and 2.5% (N=365) in the DMC. Overall, the IEWS performed significantly better than NEWS with an AUROC of 0.89 (95% confidence Intervals 0.89-0.90) versus 0.82 (0.82-0.83) in the NEED and 0.87 (0.85-0.88) versus 0.82 (0.80-0.84) at external validation. Calibration for NEWS predictions underestimated risk in older patients and overestimated risk in the youngest, while calibration improved for IEWS with a substantial reclassification of patients from low to high risk and a standardized net benefit of 5-15% in the relevant risk range for all age categories.

Conclusions The IEWS substantially improves in-hospital mortality prediction for all ED patients ≥18years.

INTRODUCTION

Early Warning Scores (EWS) are widely used prediction tools to early detect clinical deterioration of patients and trigger intensive care consultation.[9, 10, 52, 169] By aggregating points for the degree of abnormality of each vital sign, EWS provide a likelihood for mortality, which should trigger the nurse or physician to get help or to intensify treatment. These scores are widely used in many settings and they are mandatory as a standard of care in the United Kingdom.[131] The National Early warning Score (NEWS) in particular has been widely implemented and is the most frequently used score to help identify critically ill patients early.[9, 131, 170, 171]

Some limitations of the NEWS and other EWS exist. Calibration of NEWS predictions is poor with relative overestimation of risk in younger ED patients and underestimation of risk in older ED patients.[27, 42, 119] NEWS assigns 0 to 3 points for all vital signs implying that all vital signs have similar predictive value, which has been shown to be unfounded.[11, 82] Furthermore, important risk differences exist between men and women at arrival to the ED.[172, 173] Nonetheless, most studies do not test the performance of early warning scores at older age or include sex differences.[10, 30] As a result, using NEWS may cause serious disadvantages for patient care and wrong treatment or disposition decisions.

The aim of this study was to develop and externally validate an International Early Warning Score (IEWS), by recalibrating NEWS including age and sex, to improve in-hospital mortality assessment at arrival to the ED.

METHODS

Study design and setting

This international multicenter cohort study is based on existing cohorts and reporting adheres to the Transparent Reporting of a multivariable model for Individual Prognosis or Diagnosis (TRIPOD) guidelines for prognostic modelling studies.[136] The Netherlands Emergency department Evaluation Database (NEED) was used as development cohort, consisting of three hospitals in the Netherlands. The NEED is the national quality registry for Emergency Departments (EDs) in the Netherlands and contributes to the improvement of transparency and quality of ED care in the Netherlands by supplying reliable data to the participating centers (see www.stichting-need.nl).[82] Data were prospectively collected and reviewed retrospectively. Data from the three sites spanned slightly different periods: data from one tertiary center (Leiden University Medical Center) included visits between 1 January 2017 – 8 June 2019, and data from the two level II emergency centers (Medical Center Leeuwarden and Catharina Hospital Eindhoven) were from 1 January 2019 – 12 January 2020 and from 1 January 2017 – 31 December 2019, respectively.

For external validation, we used the Danish Multicenter Cohort (DMC) which has been described previously.[118, 119] These data were not only from a different setting, but also from a different period to strengthen our validation. 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).

Ethical considerations

In the Netherlands, the study was approved by the medical ethics committee of the Máxima MC on 2021, February 2 (ref number: Institutional Review Board N21.007). Under Danish law, retrospective registry studies are exempt from the need for approval by an ethics committee.[140] The study has been performed in accordance with the Helsinki declaration of 1975.

Selection of participants

All consecutive ED patients of ≥18 years were included in this study. Patients were excluded in the NEED if none or only one vital sign (systolic blood pressure (SBP), heart rate (HR), peripheral oxygen saturation (SpO_2), respiratory rate (RR) or temperature) were registered, as vital signs were considered missing not at random which prevented the possibility

for imputation (see supplementary file 1). Both studies collected data prospectively, but the DMC did so based on a prospective study design and the NEED was based on a registry. Hence, the missing data mechanisms differed for DMC (see supplementary file 2). Here, patients were excluded if neither systolic blood pressure nor pulse were recorded as these observations were missing not at random, i.e., unrelated to any of the observed variables, including outcomes.

Data collection

Demographic data were extracted from registers for both the NEED and DMC. Implausible physiological values were considered missing. Vital signs were recorded by a nurse in triage before ED treatment as described previously for the NEED,[82] and for the DMC.[118, 119]The first initial set of vital signs was registered before treatment.

NEWS aggregates seven vital signs (see table 1).[9] The NEWS was calculated for each patient (0-23points).[9] The collected Glasgow Coma Scale (GCS) was converted to an AVPU score.[119]

Table 1 The National Early Warning Score (NEWS)

Points	3	2	1	0	1	2	3
RR (/min.)	≤8		9-11	12-20		21-24	≥25
SpO2 (%)	≤91	92-93	94-95	≥96			
supplemental oxygen		Yes		No			
Temperature (°C)	≤35.0		35.1-36.0	36.1-38.0	38.1-39	≥39.1	
SBP (mmHg)	≤90	91-100	101-110	111-219			≥220
Pulse (bpm)	≤40		41-50	51-90	91-110	111-130	≥131
Level of consciousness				A			V, P, or U

RR: Respiratory Rate, $SpO_{2:}$ Peripheral oxygen saturation, SBP: Systolic blood pressure, °C: degrees Celsius, mmHg: millimeter mercury, bpm: beats per minute, A: Alert, V: Verbal, P:Pain, U: Unresponsive.

Outcome

The primary outcome was in-hospital mortality (including death in the ED). This outcome measure allowed us to compare our findings with previous studies.[119, 130] In the NEED, outcome information was registered and collected

from the minimal data set. In the DMC, information regarding mortality was collected retrospectively from the Danish Civil Registration System and the Danish National Patient Register.

Sample size estimation
See supplementary file 3.

Data analyses

Descriptive analyses
Data were presented as mean (standard deviation (SD)) if normally distributed and median (interquartile range (IQR)) if skewed.

Main statistical analyses
Predictive performance of NEWS and a recalibrated NEWS were evaluated in three age categories (18-65; 66-80; >80 years). These age categories were chosen based on previous age stratification.[82, 119] Prior to analyses, we assessed non-linearity of age in univariable logistic regression and explored non-linear terms (quadratic and restricted cubic splines) for best fit. Because patients were included if at least two vital signs were registered, missing data in the NEED were substituted by multiple imputation to reduce information bias described in supplementary file 3.[174]

For each imputation set, we calculated the NEWS. We used the vital sign categories as used in the NEWS as ordinal variables to fit the new model to prevent introducing thresholds different to those professionals are used to in current clinical practice with NEWS. We fitted the model NEWS+age+sex on the imputed data by multivariable logistic regression and, in a backwards selection approach, tested one-way and two-way interactions among predictors and found none of sufficient impact to include in the revised model. After deciding on recalibration, points were assigned and rounded to a recalibrated NEWS score based on a nomogram presentation, i.e., regression coefficients.[175] Points were rounded to nearest integer.

Predictive performance was compared in all three age categories of NEWS, recalibrated NEWS+age and the recalibrated NEWS+age+sex using Area under the Receiving Operating Characteristic (AUROC) with 95% Confidence Intervals (CI) and calibration plots. We averaged regression coefficients and intercepts across imputed sets to incorporate variance introduced by the imputation procedure. The best of the two recalibrated models was named the International Early Warning Score (IEWS).

To compare the net benefit of IEWS with NEWS, decision curves are presented.[176] This plots net-benefit at a range of risk thresholds for in-hospital mortality with the trade-off of benefit (true positive proportion) and harms (false positive proportion) on the same scale, adjusted by an appropriate exchange rate.[146] Because risk thresholds may differ by age group, separate decision curves were produced. To demonstrate how IEWS classifies patients differently than NEWS, a reclassification table was produced in which patients were allocated to low risk, medium risk or high risk subsets, stratified by outcome. In this example, we decided that the threshold from low to medium risk was two times the baseline risk and medium to high risk was three times the baseline risk (mean in-hospital mortality for patients with a NEWS <4points) per age category.

Internal and external validation
See supplementary file 3.

All analyses were performed in R statistical software (packages dplyr (v1.0.7;2021), rms (v6.2;2021), mice(v45;2011). A p-value <0.05 was considered as statistically significant.

RESULTS

Patient characteristics
In total, 95,553 patients could be included for analyses from the NEED with mean age 60.1 years (SD 19.4) and 50.3% male patients. Patient characteristics are described in supplementary file 4. Excluded patients had lower in-hospital mortality and fewer Intensive Care Unit (ICU) admissions than the included patients (see supplementary file 1). For external validation, a total of 14,809 patients were included. They had a mean age of 63 (SD 20) years, 51.9% were male. Patient characteristics in DMC were comparable to the NEED (supplementary file 5).

Main results
Age was used as a linear spline with no age effect assumed below 40 years based on its fit and association with mortality (supplementary file 6). A nomogram for the recalibrated NEWS plus age and sex was presented (supplementary file 7).

Based on the nomogram, points were assigned for a recalibrated NEWS+age score and a recalibrated NEWS+age+sex score resulting in a new risk score (table 2). Based on the calibration plots for NEED data (Figure 1) and for the DMC (Figure 2), the NEWS+age+sex was chosen as the best fit because

calibration improved visually and according to the slope and intercept in the relevant risk range for all age categories while discrimination was not affected by adding sex. The NEWS+age+sex model was therefore proposed as the IEWS. Flexible calibration curves are shown in supplementary file 8 and 9 for NEED data and DMC. Whereas the NEWS showed substantial underestimation of risk in older patients and overestimation of risk in younger patients, calibration for IEWS improved in both the development and validation cohort (Figures 1 and 2).

Table 2 The International Early Warning Score (IEWS)

Points	5	3	2	1	0	1	2	3	5
RR (/min.)					0-20		21-24	≥25	
SpO2 (%)		≤91		92-95	≥96				
supplemental oxygen				Yes	No				
Temperature (°C)	≤35.0	35.1-36.0			≥36.1				
SBP (mmHg)			≤90	91-110	111-219		≥220		
Pulse (bpm)				≤50	51-90	91-110	≥111		
Level of consciousness					A				V,P, or U
Sex				Male	Female				
Age, years					0-40	41-50	51-60	61-65 (66-75= 4points)	76-80=5 (81-90= 6points) (91-100= 7points)

RR: Respiratory Rate, SpO_2: Peripheral oxygen saturation, SBP: Systolic blood pressure, °C: degrees Celsius, mmHg: millimeter mercury, bpm: beats per minute, A: Alert, V: Verbal, P:Pain, U: Unresponsive.
The International Early Warning Score (IEWS) is a recalibrated model of the National Early Warning Score (NEWS) extended with age and sex.

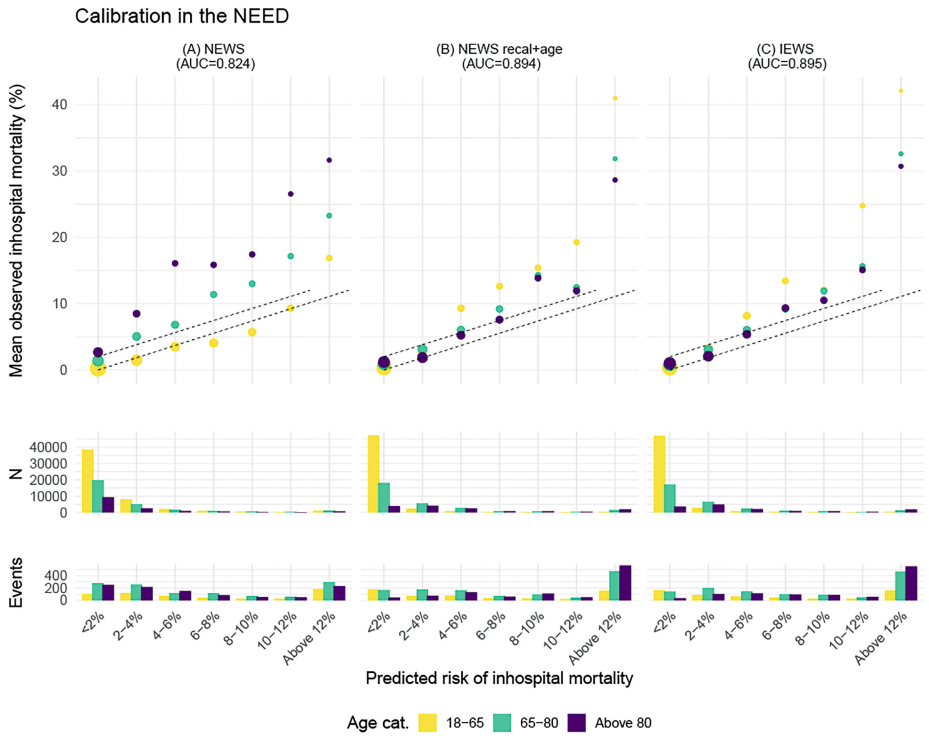


Figure 1 Internal calibration plots for the NEED (Netherlands Emergency department Evaluation Database) for (A) the National Early Warning Score (NEWS), (B) a recalibrated NEWS + age and (C) a recalibrated NEWS+age+sex (the International Early Warning Score (IEWS)). The predicted in-hospital mortality was categorized in steps of 2% in the relevant risk range. Calibration was assessed in three different age categories (18-65y, 66-80y,>80y). The 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 based on the inverse of the standard deviation. Below the calibration figures are the distribution of patients and outcomes presented for all three scores in histograms.

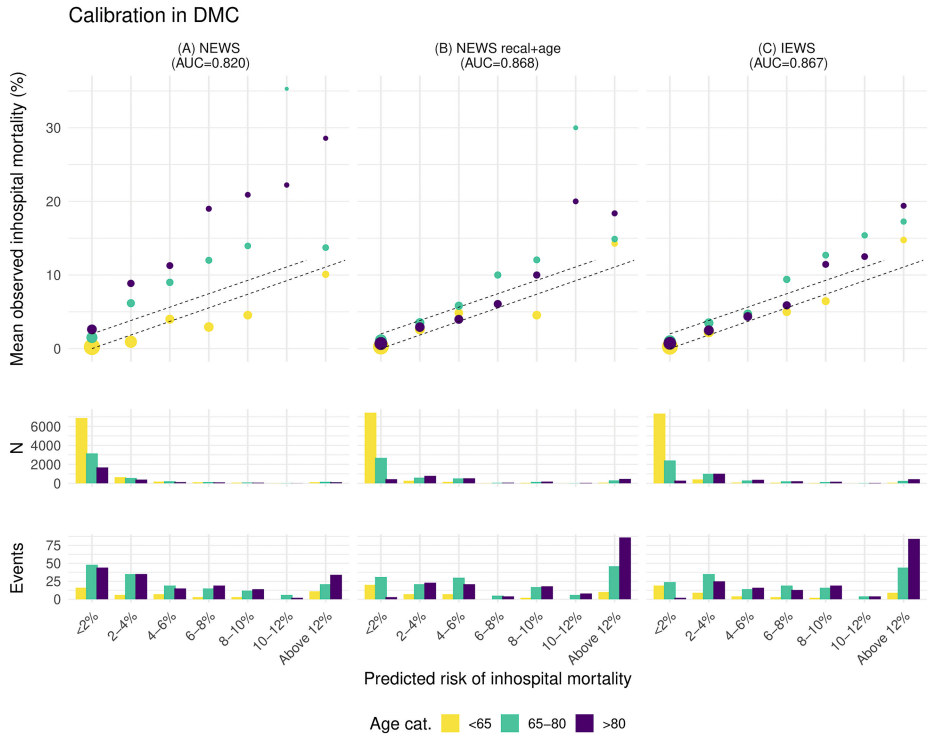


Figure 2 External calibration plots for the Danish Multicenter Cohort (DMC) for (A) the National Early Warning Score (NEWS), (B) a recalibrated NEWS + age and (C) a recalibrated NEWS+age+sex (the International Early Warning Score (IEWS)). The predicted inhospital mortality was categorized in steps of 2% in the relevant risk range. Calibration was assessed in three different age categories (18-65y, 66-80y,>80y). The 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 based on the inverse of the standard deviation. Below the calibration figures are the distribution of patients and outcomes presented for all three scores in histograms.

Overall, AUROC improved substantially for IEWS with 0.89 (95% CI 0.89-0.90) compared to NEWS 0.82 (95% CI 0.82-0.83 in the NEED and in the DMC with AUROC for IEWS 0.87 (95% CI 0.85-0.88) compared to NEWS 0.82 (95% CI 0.75-0.89). For most age categories, discrimination improved substantially (table 3). Internal validation showed good performance of IEWS (Supplementary file 10). Split sample analyses based on hospital location showed similar results (Supplementary file 11).

Table 3 Calibration and discrimination for NEWS and IEWS in the development and validation cohort

	Calibration		Discrimination	
Age groups	**Intercept**	**Slope**	**AUROC**	**95% CI**
Development cohort				
NEWS for in-hospital mortality in the NEED				
18-65y	-0.68	1.32	0.87	0.85-0.88
66-80y	0.38	0.97	0.80	0.79-0.81
>80y	0.97	0.91	0.78	0.77-0.80
Overall	0.18	1.09	0.82	0.82-0.83
IEWS for in-hospital mortality in the NEED				
18-65y	0.15	1.47	0.92	0.90-0.93
66-80y	0.21	1.23	0.85	0.84-0.86
>80y	0.16	1.18	0.83	0.82-0.85
Overall	0.18	1.24	0.89	0.89-0.90
Validation Cohort				
NEWS for in-hospital mortality in the DMC				
18-65y	-1.05	1.09	0.82	0.75-0.89
66-80y	0.43	0.82	0.78	0.74-0.82
>80y	0.94	0.84	0.78	0.74-0.81
Overall	0.20	0.98	0.82	0.80-0.84
IEWS for in-hospital mortality in the DMC				
18-65y	-0.52	1.05	0.86	0.80-0.91
66-80y	-0.04	0.88	0.80	0.76-0.83
>80y	-0.25	0.83	0.77	0.73-0.81
Overall	-0.21	0.94	0.87	0.85-0.88

NEWS: National Early Warning Score, IEWS: International Early Warning score, AUROC: Area under the Receiving Operating Curve, 95% CI: 95percent Confidence Intervals, NEED: Netherlands Emergency Department Evaluation Database, DK: Danish Multicenter Cohort.
The International Early Warning Score(IEWS) is a recalibrated model of the NEWS including the additional variables age and sex.

Decision curve analyses showed for each age category a standardized Net Benefit of 5-15% in the relevant risk range of 1% to 15% (figure 3). As an example, in a population with approximately 24 in-hospital deaths per 1000 patients, for a decision threshold of 5% in-hospital mortality risk, the IEWS would identify 42% additional true deaths (Standardized net benefit at a threshold of 5% for IEWS in figure 3), without increasing the number of false positive predictions compared to not using any model. Compared to using NEWS, the IEWS would identify 15% additional true deaths without increasing the number of false positive predictions.

To give a better insight in the benefit of using IEWS compared to NEWS, a reclassification table is presented in Supplementary file 12.

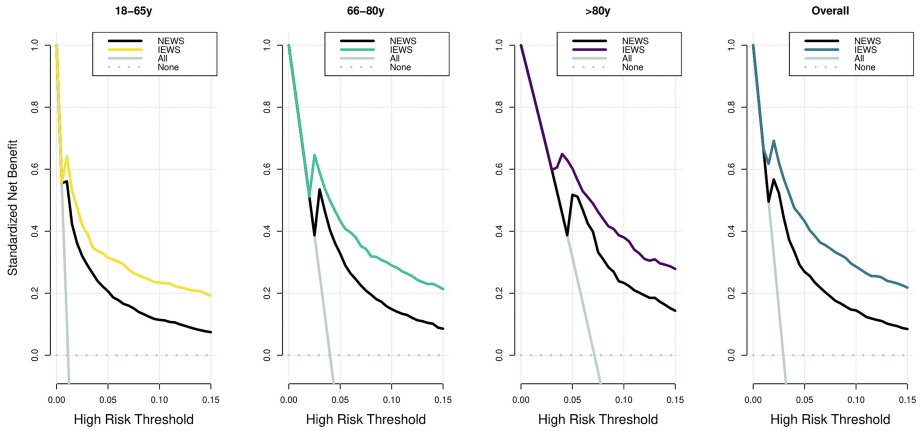


Figure 3 The decision curve analyses show for each age category a standardized Net Benefit of 5-15% in the relevant risk range of 1% to 15%. One physician may be more defensive and use, for example, 5% as a threshold for additional interventions (i.e. Intensive Care consultation or broad spectrum antibiotics) for patients >80years and older whereas another physician may choose to use a threshold of 10% for older patients. With both thresholds, the IEWS performs better than the NEWS, with 10-15% net more true positives (i.e. identified more patients who died corrected for the number of false positive classifications).

DISCUSSION

This large international multicenter cohort study shows that the IEWS, a recalibrated model based on NEWS including age and sex, performs significantly better compared to the widely adopted NEWS for the prediction of in-hospital mortality in ED patients of all age categories in a development and external validation cohort.

Most early warning scores have methodological weaknesses that could have detrimental effects on patient care if used in clinical practice.[130, 177] For example, NEWS, based on the VitalPAC early warning score (ViEWS), did not include age because AUROC only slightly improved after including age.[30, 13]

[3] However, calibration, a key aspect of prediction model performance,[130, 136] has not been assessed and age was used as a dichotomous variable below or above 65 years instead as continuous predictor. Furthermore, points for vital sign disturbances were allocated based on clinical consensus rather than on a statistical approach.[10, 30]

Our results are in line with several studies which have demonstrated that including age to an early warning score improved predictive performance substantially.[42, 133, 135, 151] However, none of these studies followed the recommended steps for development and validation of prediction models neither have they shown a classification in low to high risk.[178, 179] Our decision curve analysis and reclassification table demonstrate that for both younger and older patients the IEWS has considerable incremental value with more young patients correctly classified as low risk, and more importantly, more older patients correctly classified as high risk for in-hospital mortality. Previous studies have shown that predictive performance only improved for younger patients using an age-specific early warning score on a composite outcome of mortality, cardiac arrest and ICU admission compared to NEWS.[166, 180, 181] However, the modeling approach was very different from ours. Points were assigned to vital signs based on their distribution rather than on regression coefficients as recommended for prediction modeling.[130, 174, 178] This may have caused the age-specific model to underperform in older age. Our group has demonstrated previously that the addition of age to NEWS without recalibration of the physiological variables already improved predictive performance for in-hospital mortality.[119]

Early warning scores are designed for prognostication and can be used as early as in the ED and add to the clinical evaluation of a patient's disease severity.[182] While clinical evaluation may vary among physicians depending on years of experience,[183] the IEWS provides a numerical mortality risk (a percentage) that hypothetically may help with clinical decision-making.

For an easily adopted and implemented early warning score it is essential that the variables in the score are easily measured, readily available and strong predictors of the primary outcome.[176, 178] The physiological variables used in NEWS meet all these requirements.[11, 82] In addition, age and sex exhibit the same qualities and are predictors for in-hospital mortality.[119, 172]

Other variables have been proposed to use in early warning scores, such as biomarkers or frailty measures.[73, 156, 184-186] For frailty, only four out of 60 frailty scores could be measured in less than one minute using vignettes.[187] Though, in clinical practice, it may be difficult and not reliable to assess frailty, for example if the level of consciousness is altered and no history is available. Other variables such as biomarkers are not readily available or easily repeated without high costs. For these reasons, we have only evaluated age and sex as additional variables to the seven predictors of NEWS which both met the criteria for reliable predictors and are always known or can at least be estimated precisely.[130, 178] In patients who received prehospital treatments from paramedics or medical emergency services, the physiological variables may already be improved at arrival to the ED and thus the risk may be underestimated by using the NEWS or IEWS, a phenomenon called lead-time bias in literature.[188] Prehospital treatments have not been considered in the model. However, the IEWS still performs better overall than NEWS also in the ED.

The present study has several strengths. We adhered to the TRIPOD guidelines and followed steps recommended for the development of prediction models.[130, 136, 174, 178]. We used a large sample size in relation to the number of predictors for both development and validation and validated our findings externally in a different European country in a different time-period. The IEWS is clinically useful in the relevant risk range for each age category. Further validation is desired to assess generalizability of the proposed IEWS across multiple settings.[189, 190] Other limitations need to be considered. First, a risk of selection bias may be present as we excluded patients in whom less than two vital signs were registered. However, these patients were at very low risk of mortality (for example wounds and fractures) or at very high risk (cardiac arrest) and therefore these patients would have been recognized as low or high risk also without an early warning score. As recommended, we used multiple imputation to prevent information bias so we could include as many patients as possible.[130] Notably, around 90% of AVPU values were missing in the development cohort. However, missingness was clearly related to outcomes and other measured variables (Supplementary file 13), IEWS worked very well in the external validation cohort with a very low missingness of the AVPU variable. Hence, the bias incurred by imputing AVPU is likely negligible, despite a high proportion of missingness.[191] Secondly, in-hospital mortality was chosen as the primary outcome. A time horizon of a few days only is recommended for early warning scores.[130] Nevertheless, the time till patients died in-hospital in our data was short with a median of 4days (IQR 1-9), which allowed us to compare our results with previous studies and assess deterioration of patients.[9, 131, 192] Thirdly, the physiological

variables have been categorized based on the NEWS because physicians are used to work with these thresholds in clinical practice. However, it has been recommended to avoid categorizing predictors in the statistical analysis. For this reason, we repeated our analysis using restricted cubic splines for each physiological variable and presented a nomogram (Supplementary file 14). Using this nomogram would have resulted in similar distribution of points as in the IEWS after rounding. Thus, categorization of variables did not lead to poor modelling. Nonetheless, using different points for each physiological variable may lead to calculation errors as physicians are used to using the NEWS. This could be overcome by calculating the score electronically. Lastly, the NEWS2, a modification of the original NEWS, has not been evaluated in this study for several reasons. First, mortality assessment did not improve using NEWS2 compared to NEWS in a previous large study.[193] Secondly, the two major updates introduced in NEWS2 were separate thresholds for saturation in patients with hypercapnic failure and the addition of confusion in consciousness scale. We did not record confusion, which makes it impossible to use the NEWS2 consciousness scale. Additionally, information about current or previous hypercapnic failure is often not available at arrival or requires arterial blood gas. We therefore bases the IEWS on the foundation laid out in NEWS rather than NEWS2. Comparing the IEWS with other widely adopted EWS, such as the Modified Early Warning Score (MEWS), would have resulted in similar results, as the design of MEWS was neither based on a statistical approach, nor it includes age or sex.[135]

In summary, this large international multicenter cohort study shows that the IEWS performs substantially better than the widely adopted NEWS for predicting mortality in ED patients of all age categories in a development and external validation cohort. Future studies should investigate further evidence for predictive validity and assess whether implementation of IEWS in the ED leads to lower adverse events compared to not using an early warning score or using NEWS.

SUPPLEMENTARY FILE 1

Patient characteristics of excluded patients: Patient characteristics of both included (0-3 missing vitals) and excluded patients (4-5missing vitals) are presented in the table for comparison. Patients with four or five missing vital signs (Systolic blood pressure, temperature, heart rate, peripheral oxygen saturation or respiratory rate) were excluded from the analyses, because vital signs were considered missing not at random in these patients. Excluded patients are described in the table.

Excluded patients had lower in-hospital mortality, less ICU admissions and lower urgency of triage compared to included patients.

		Included patients Number of missing vital signs			**Excluded patients**	
NEED cohort	**0 (N=50378)**	**1 (N=29174)**	**2 (N=10383)**	**3 (N=5618)**	**4 (N=5843)**	**5 (N=47393)**
In-hospital mortality						
died	1406 (2.8%)	614 (2.1%)	215 (2.1%)	79 (1.4%)	57 (1.0%)	478 (1.0%)
Missing	681 (1.4%)	437 (1.5%)	210 (2.0%)	159 (2.8%)	148 (2.5%)	519 (1.1%)
ICU admission						
ICU admission	821 (1.6%)	549 (1.9%)	143 (1.4%)	48 (0.9%)	31 (0.5%)	310 (0.7%)
No ICU admission	49195 (97.7%)	28267 (96.9%)	9965 (96.0%)	4770 (84.9%)	5145 (88.1%)	39976 (84.4%)
Missing	362 (0.7%)	358 (1.2%)	275 (2.6%)	800 (14.2%)	667 (11.4%)	7107 (15.0%)
Triage category*						
immediate	3376 (6.7%)	1370 (4.7%)	396 (3.8%)	110 (2.0%)	86 (1.5%)	632 (1.3%)
very urgent	17015 (33.8%)	7761 (26.6%)	2218 (21.4%)	665 (11.8%)	516 (8.8%)	4913 (10.4%)
urgent	20768 (41.2%)	12855 (44.1%)	4731 (45.6%)	2927 (52.1%)	2515 (43.0%)	16502 (34.8%)
non-urgent	8552 (17.0%)	6600 (22.6%)	2825 (27.2%)	1835 (32.7%)	2595 (44.4%)	20845 (44.0%)
Missing	667 (1.3%)	588 (2.0%)	213 (2.1%)	81 (1.4%)	131 (2.2%)	4501 (9.5%)

*Triage category according to the Manchester Triage System or Dutch Triage Standard.

SUPPLEMENTARY FILE 2

Patient characteristics of excluded patients in DMC: Patients were excluded if neither systolic blood pressure nor pulse were recorded as these observations were missing not at random, i.e., unrelated to any of the observed variables, including outcomes.

DMC Cohort	**Included N=14809**						**Excluded N=2039**				**Pearson χ^2**
	Number of missing vital signs (frequency)										
	0 (9132)	1 (4075)	2 (1357)	3 (217)	4 (21)	5 (7)	3 (4)	4 (7)	5 (201)	6 (1827)	
Inhospital mortality, N (%)	215 (2.4)	105 (2.6)	36 (2.7)	8 (3.7)	1 (4.8)	0 (0)	1 (25.0)	0 (0)	7 (3.5)	50 (2.7)	
Subtotal	365 (2.5)						59 (2.8)				χ^2=1.06, P=0.304
ICU admission, N (%)	220 (2.4)	142 (3.5)	46 (3.4)	8 (3.7)	1 (4.8)	0 (0)	1 (25.0)	1 (14.3)	7 (3.5)	51 (2.8)	
Subtotal	417 (2.8)						60 (2.9)				χ^2=0.1, P=0.746

Missing vital signs include respiratory rate, heart rate, systolic blood pressure, peripheral oxygen saturation, temperature and avpu.

SUPPLEMENTARY FILE 3

Sample size estimation

The NEED contained 148,828 ED visits of patients ≥18 years. We estimated that in ~60% of the ED visits at least two or more vital signs were registered resulting in ~90,000 ED visits which could be used for the analyses with approximately 2300 events (in-hospital mortality). This number is more than sufficient for reliable analyses.[178] Numbers were also large in age-based subgroups. For external validation a minimum of in total 200 events is recommended.[178] The DMC contains approximately 350 events which should be sufficient for external validation.[119]

Multiple imputation procedure

Because patients were included if at least two vital signs were registered, missing data in the NEED were substituted by multiple imputation to reduce information bias.[174] We used the chained equations procedure, after imputation was deemed feasible based on patterns of missingness.[179, 194] For a better multiple imputation procedure, we also used triage category (non-urgent, urgent, very urgent, immediate; according to the Manchester Triage System or Dutch Triage Standard), urea, leukocytes, and fluid administration (0, 0-500ml, >500ml) as a predictor in the imputation procedure. Outcome was imputed if missing. Imputation parameters for DMC have been described in detail previously.[119] We obtained 20 estimates of the missing vital signs for each patient with five iterations each. We checked for collinearity and convergence during the imputation procedure.

Internal and external validation

For internal validation, a bootstrapping was performed with 200 repetitions on the imputed data and the overall AUROC and calibration were presented.[195] Also, a non-randomly split sample analysis was performed based on hospital location in the NEED cohort.

The DMC was used for external validation. Predictive performance in terms of discrimination and calibration was assessed and based on the average estimate of risk. A net benefit curve was produced.

SUPPLEMENTARY FILE 4

Patient characteristics of the NEED, the Development cohort

NEED cohort	18-65years (N=51573)	66-80years (N=29591)	>80years (N=14389)	All (N=95553)
Age, years				
Mean (SD)	45.7 (14.0)	72.9 (4.21)	85.8 (3.87)	60.1 (19.4)
Sex				
male	25424 (49.3%)	16363 (55.3%)	6316 (43.9%)	48103 (50.3%)
Systolic Blood Pressure (mmHg)				
Mean (SD)	128 (28.2)	137 (33.2)	142 (34.3)	133 (31.3)
Missing	4380 (8.5%)	1525 (5.2%)	559 (3.9%)	6464 (6.8%)
Heart rate (bpm)				
Mean (SD)	87.0 (20.7)	85.7 (21.6)	83.4 (20.8)	86.0 (21.0)
Missing	7037 (13.6%)	2647 (8.9%)	1232 (8.6%)	10916 (11.4%)
Respiratory Rate (/min.)				
Median [IQR]	16.0 [14.0-19.0]	17.0 [15.0-21.0]	18.0 [15.0-22.0]	17.0 [14.0-20.0]
Missing	16800 (32.6%)	7103 (24.0%)	3226 (22.4%)	27129 (28.4%)
Peripheral oxygen saturation (%)				
Mean (SD)	97.8 (2.96)	96.4 (3.81)	96.0 (3.84)	97.1 (3.47)
Missing	2443 (4.7%)	1397 (4.7%)	702 (4.9%)	4542 (4.8%)
Level of consciousness				
Alert	4311 (8.4%)	3290 (11.1%)	1840 (12.8%)	9441 (9.9%)
Verbal	178 (0.3%)	118 (0.4%)	114 (0.8%)	410 (0.4%)
Pain	109 (0.2%)	44 (0.1%)	36 (0.3%)	189 (0.2%)
Unresponsive	57 (0.1%)	42 (0.1%)	15 (0.1%)	114 (0.1%)
Missing	46918 (91.0%)	26097 (88.2%)	12384 (86.1%)	85399 (89.4%)
Temperature (degrees Celcius)				
Median [IQR]	37.0 [36.5-37.4]	36.8 [36.5-37.4]	36.8 [36.4-37.3]	36.9 [36.5-37.4]
Missing	9820 (19.0%)	5161 (17.4%)	2762 (19.2%)	17743 (18.6%)

NEED cohort	18-65years (N=51573)	66-80years (N=29591)	>80years (N=14389)	All (N=95553)
Supplemental Oxygen				
yes	1421 (2.8%)	2130 (7.2%)	1449 (10.1%)	5000 (5.2%)
Missing	15931 (30.9%)	6984 (23.6%)	2672 (18.6%)	25587 (26.8%)
Fluid administration (ml)				
0cc	24458 (47.4%)	16065 (54.3%)	7807 (54.3%)	48330 (50.6%)
0-500cc	4593 (8.9%)	2850 (9.6%)	1496 (10.4%)	8939 (9.4%)
>500cc	5248 (10.2%)	3231 (10.9%)	1377 (9.6%)	9856 (10.3%)
Missing	17274 (33.5%)	7445 (25.2%)	3709 (25.8%)	28428 (29.8%)
Triage category, N (%)	*			
non-urgent	11076 (21.5%)	5683 (19.2%)	3053 (21.2%)	19812 (20.7%)
urgent	22703 (44.0%)	12510 (42.3%)	6068 (42.2%)	41281 (43.2%)
very urgent	14413 (27.9%)	9071 (30.7%)	4175 (29.0%)	27659 (28.9%)
immediate	2541 (4.9%)	1871 (6.3%)	840 (5.8%)	5252 (5.5%)
Missing	840 (1.6%)	456 (1.5%)	253 (1.8%)	1549 (1.6%)
Urea (mmol/L)				
Median [IQR]	4.9 [3.8-6.2]	6.7 [5.3-9.0]	8.2 [6.2-11.4]	5.8 [4.4-7.9]
Missing	11273 (21.9%)	4435 (15.0%)	2375 (16.5%)	18083 (18.9%)
Leukocytes (*10^9)				
Median [Min, Max]	9.2 [7.0-12.0]	9.0 [6.9-12.1]	9.2 [7.1-12.2]	9.1 [7.0-12.1]
Missing	10667 (20.7%)	4001 (13.5%)	1921 (13.4%)	16589 (17.4%)
In-hospital mortality, N(%)				
Died	488 (0.9%)	981 (3.3%)	845 (5.9%)	2314 (2.4%)
Missing	608 (1.2%)	554 (1.9%)	325 (2.3%)	1487 (1.6%)
ICU admission, N (%)				

NEED cohort	18-65years (N=51573)	66-80years (N=29591)	>80years (N=14389)	All (N=95553)
ICU admission	828 (1.6%)	582 (2.0%)	151 (1.0%)	1561 (1.6%)
Missing	1385 (2.7%)	319 (1.1%)	91 (0.6%)	1795 (1.9%)

SD: Standard deviation, IQR: Interquartile Range, ICU: Intensive Care Unit
*Triage category according to the Manchester Triage System or Dutch Triage Standard.

SUPPLEMENTARY FILE 5

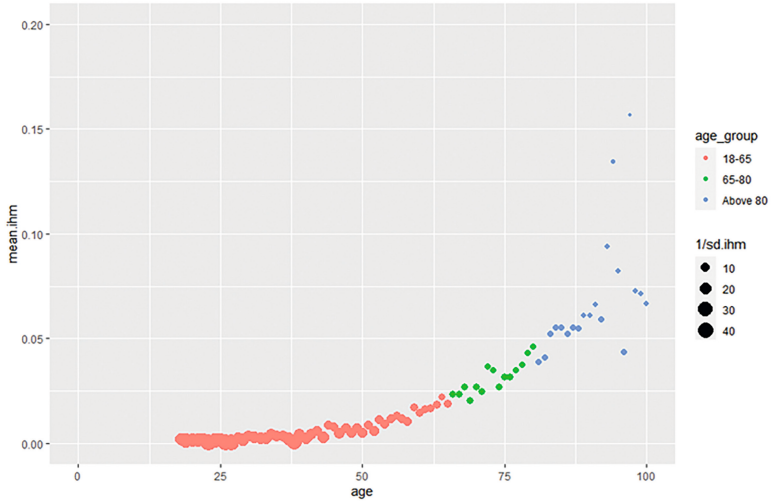
Patient characteristics of the Danish Multicenter Cohort (DMC) used for external validation.

DMC cohort	<65 (N=7990)	65-80 (N=4316)	>80 (N=2503)	All (N=14809)
Age				
Mean (SD)	44.9 (13.9)	73.0 (4.4)	86.1 (4.1)	60.0 (20.0)
Gender				
male	4,186 (52.4%)	2,146 (49.7%)	1,352 (54.0%)	7,684 (51.9%)
Systolic blood pressure (mmHg)				
Mean (SD)	135 (21.6)	140 (26.2)	140 (27.4)	137 (24.1)
Missing	21 (0.3%)	11 (0.3%)	12 (0.5%)	44 (0.3%)
Heart Rate (bpm)				
Mean (SD)	84.8 (19.6)	84.6 (20.3)	81.6 (19.6)	84.2 (19.8)
Missing	54 (0.7%)	40 (0.9%)	29 (1.2%)	123 (0.8%)
Respiratory Rate (/min)				
Median [IQR]	16 [5]	16 [5]	18 [8]	16 [6]
Missing	2,340 (29.3%)	1,073 (24.9%)	624 (24.9%)	4,037 (27.3%)
Peripheral oxygen saturation (%)				
Mean (SD)	97.1 (3.1)	95.4 (4.2)	94.9 (4.7)	96.2 (3.8)
Missing	343 (4.3%)	159 (3.7%)	135 (5.4%)	637 (4.3%)
Level of consciousness				
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	939 (11.8%)	487 (11.3%)	350 (14.0%)	1,776 (12.0%)
Temperature (Degrees Celsius)				
Median [IQR]	37.0 [0.8]	36.9 [0.9]	36.9 [0.9]	37.0 [0.8]
Missing	526 (6.6%)	234 (5.4%)	182 (7.3%)	942 (6.4%)

DMC cohort	<65 (N=7990)	65-80 (N=4316)	>80 (N=2503)	All (N=14809)
Supplementary oxygen				
Yes	353 (4.4%)	427 (9.9%)	309 (12.3%)	1,089 (7.4%)
In-hospital mortality, N (%)				
Died	46 (0.6%)	156 (3.6%)	163 (6.5%)	365 (2.5%)
ICU admission, N (%)				
ICU admission	197 (2.5%)	144 (3.3%)	76 (3.0%)	417 (2.8%)

SUPPLEMENTARY FILE 6

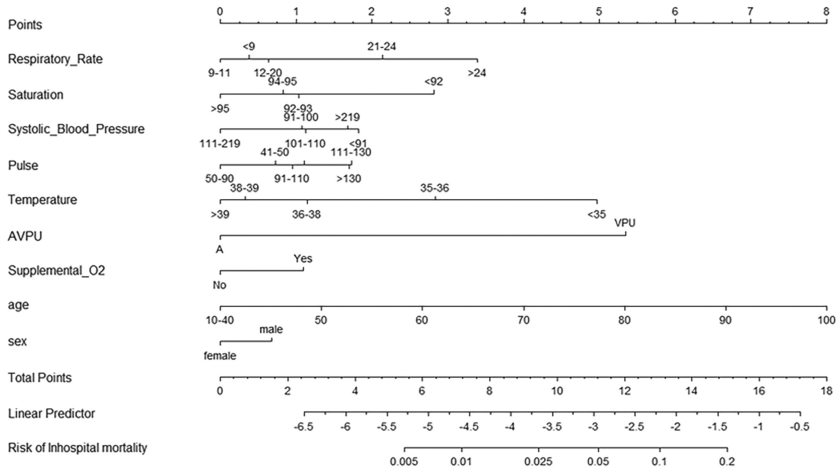
The association between age and in-hospital mortality



Ihm: in-hospital mortality
Included patients from the NEED (N=95,553). Age was used as a linear predictor in the prediction model above 40years old. The size of the dots indicates the precision of the estimate for observed in-hospital mortality and is based on the inverse of the standard deviation. The larger, the higher the precision.

SUPPLEMENTARY FILE 7

Nomogram for a recalibrated National Early Warning Score +age+sex



All predictors were used as categorized variables as used in the original National Early Warning Score (NEWS). A nomogram was developed based on the regression coefficients of all predictors (see below). A nomogram is a graphical presentation of model fit, allowing points to be awarded on a scale that is proportional to the log-odds. Points were rounded to nearest integer.
A nomogram fitted on the primary outcome (in-hospital mortality). As can be seen in the nomogram, age was the strongest predictor for mortality followed by AVPU. Male sex was associated with increased mortality. The nomogram for a recalibrated NEWS score with age and without sex was comparable to the recalibrated NEWS with age and with sex.

Regression coefficients and odds ratio's of multivariable logistic regression

Variables	Regression coefficients	Odds ratio's (95% CI)	Variables	Regression coefficients	Odds Ratio's (95% CI)
Intercept	-1,45	-	Pulse (<41bpm)	-0,22	1.5 (0.9-2.4)
RR (<9/min.)	0,4	0.84 (0.35-2.0)	Pulse (41-50bpm)	0,31	1.4 (1.0-1.9)
RR (9-11/min.)	-0,19	0.82 (0.54-1.6)	Pulse (51-90bpm)	0	Ref.
RR (12-20/min.)	0	Ref.	Pulse (91-110bpm)	-0,84	1.5 (1.3-1.6)
RR (21-24/min.)	1,01	1.87 (1.6-2.2)	Pulse (110-130bpm)	-0,23	2.0 (1.7-2.3)
RR (>24/min.)	1,91	3.1 (2.7-3.6)	Pulse(>130bpm)	-0,83	2.0 (1.6-2.5)
SPO2(>95%)	0	Ref.	Temp(<35dgr)	0,14	4.7 (3.4-6.7)
SpO2(94-95%)	0,35	1.4 (1.2-1.6)	Temp(35-36dgr)	0,71	2.0 (1.7-2.4)
SPO2(92-93%)	-0,29	1.5 (1.3-1.8)	Temp(36-38dgr)	0	Ref.
SPO2(<92%)	0,083	3.1 (2.7-3.6)	Temp(38-39dgr)	-2,44	0.72 (0.60-0.87)
SBP (>219mmHg)	-0,98	2.1 (1.5-3.1)	Temp(>39dgr)	-3,3	0.62 (0.48-0.80)
SBP(111-219mmHg)	0	Ref.	Alert	0	Ref.
SBP (101-110mmHg)	0,43	1.5 (1.3-1.8)	VPU	0,72	8.8 (7.4-10.5)
SBP(91-100mmHg)	-0,43	1.5 (1.3-1.8)	O2 (No)	0	Ref.
SBP(<91mmHg)	-0,58	2.1 (1.8-2.4)	O2 (yes)	0,23	1.6 (1.4-1.8)
Age (counting from 63)	-0,95	4.5 (4.1-5.0) /28y	Sex (female)	-0,29	0.74 (0.68-0.82)

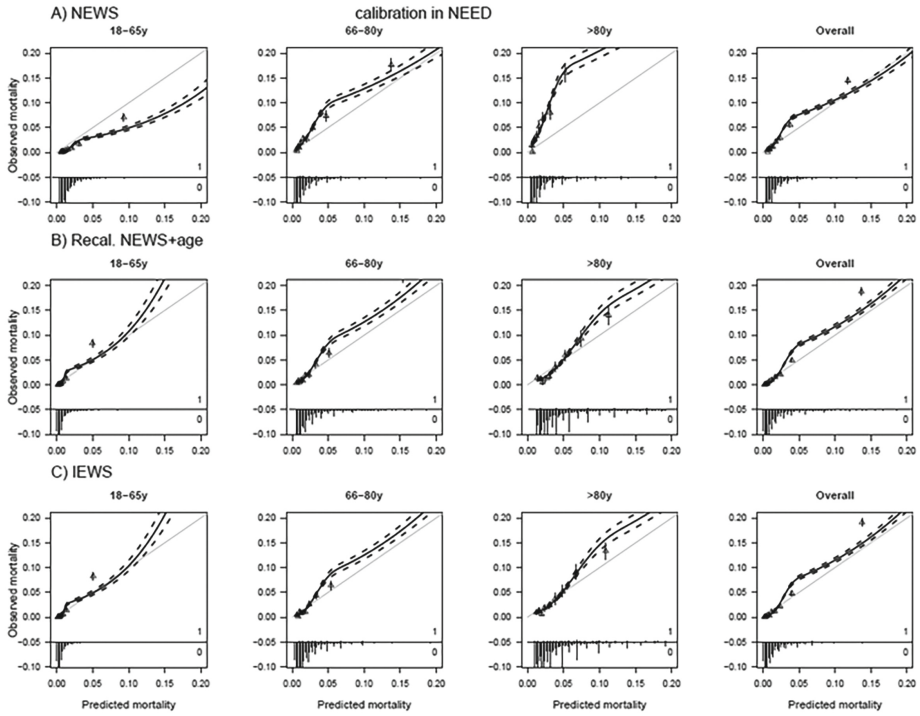
Function

function(Rr2 = 0,Sat = 0,Sbp2 = 0,Pulse2 = 0,Temp2 = 0,Avpu = 0,O2 = 0,age = 63,sex = “male”) {Rr2 <- ordered(Rr2);Sat <- ordered(Sat);Sbp2 <- ordered(Sbp2);Pulse2 <- ordered(Pulse2);Temp2 <- ordered(Temp2);age <- Function(age2)(age);

-1.4534925-0.19432847*(RR== 9-11) +1.0141945*(RR==21-24) +0.40291742*(RR<=9) +1.9103567*(RR>=24)+0.34889429*(SPO2=94-95)-0.28768878*(Sat==92-93) +0.082739612*(Sat<92) +0.43450926*(SBP=101-110)-0.43269485*(Sbp=91-100)-0.57997328*(Sbp<91)-0.97877706*(Sbp>219)+0.30637857*Pulse(41-50)-0.21608917*(Pulse<41)-0.233718-59*(Pulse= 110-130)-0.8381017*(Pulse=91-110)-0.83097791*(Pulse>130)+0.7070425*Temp(35-36)+0.14159026*(Temp<35)-2.4456138*(Temp2=38-39)-3.2986931*(Temp>39)+0.72456848*vpu+0.23021481*O2(yes)-0.94793297*age-0.29486964*(sex==”female”) }

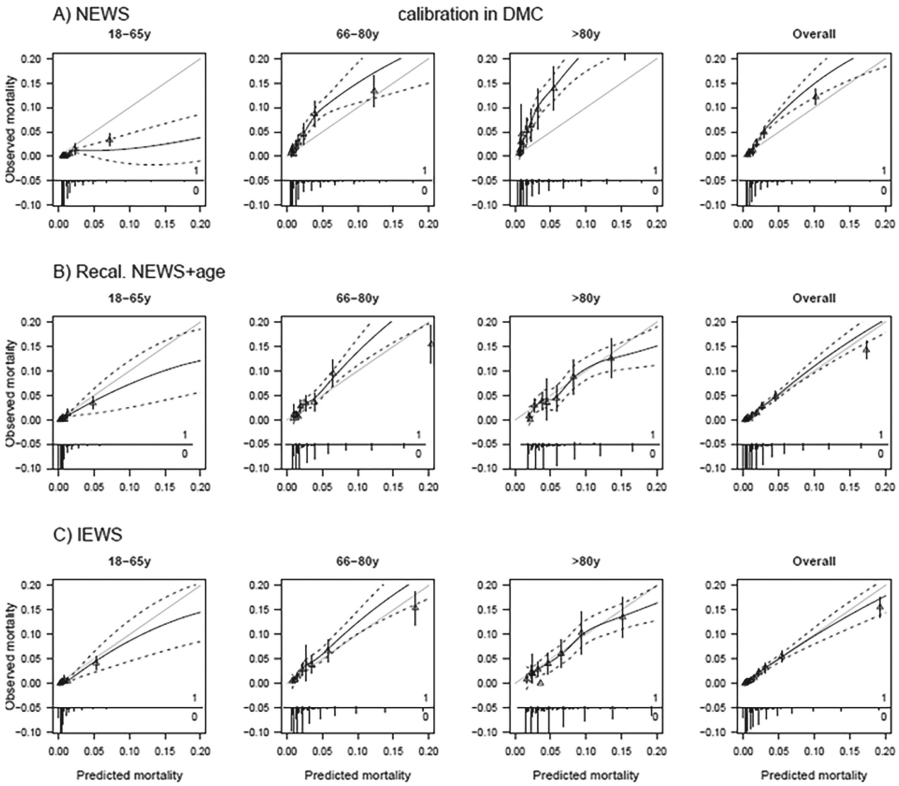
SUPPLEMENTARY FILE 8

Flexible calibration plots in the NEED



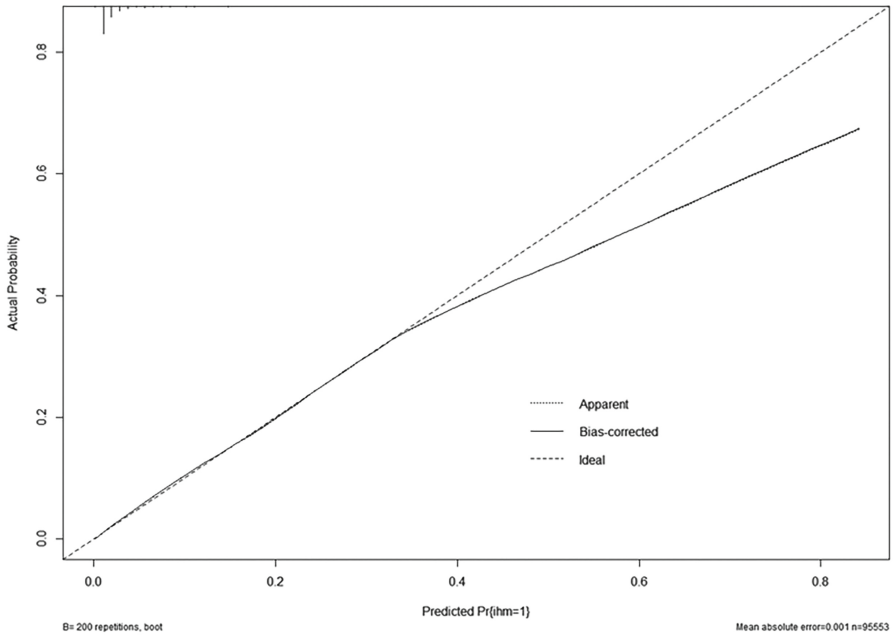
SUPPLEMENTARY FILE 9

Flexible calibration plots in the DMC



SUPPLEMENTARY FILE 10

Internal validation**:** Internal validation using bootstrap with 200 random sample repetitions on the NEED.



AUROC overall = 0.87
Calibration plot. Notes: Apparent refers to apparent performance for calibration; bias-corrected refers to optimism-corrected in internal validation; bootstrap=200.

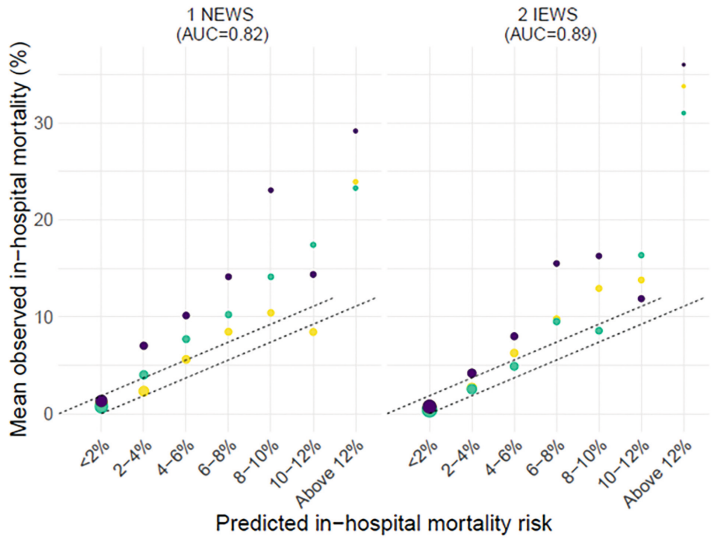
SUPPLEMENTARY FILE 11

Split sample analysis based on hospital location in the NEED.

	Discrimination	
Age groups	**AUROC**	**95% CI**
NEWS for in-hospital mortality		
Hospital 1 (tertiary care center)	0.80	0.78-0.82
Hospital 2	0.86	0.85-0.87
Hospital 3	0.82	0.81-0.84
IEWS for in-hospital mortality		
Hospital 1 (tertiary care center)	0.87	0.86-0.89
Hospital 2	0.90	0.90-0.91
Hospital 3	0.87	0.86-0.89

NEWS: National Early Warning Score, IEWS: International Early Warning score, AUROC: Area under the Receiving Operating Curve, 95% CI: 95percent Confidence Intervals, NEED: Netherlands Emergency Department Evaluation Database.
The International Early Warning Score is a recalibrated model of the NEWS including age and sex as extra variables.

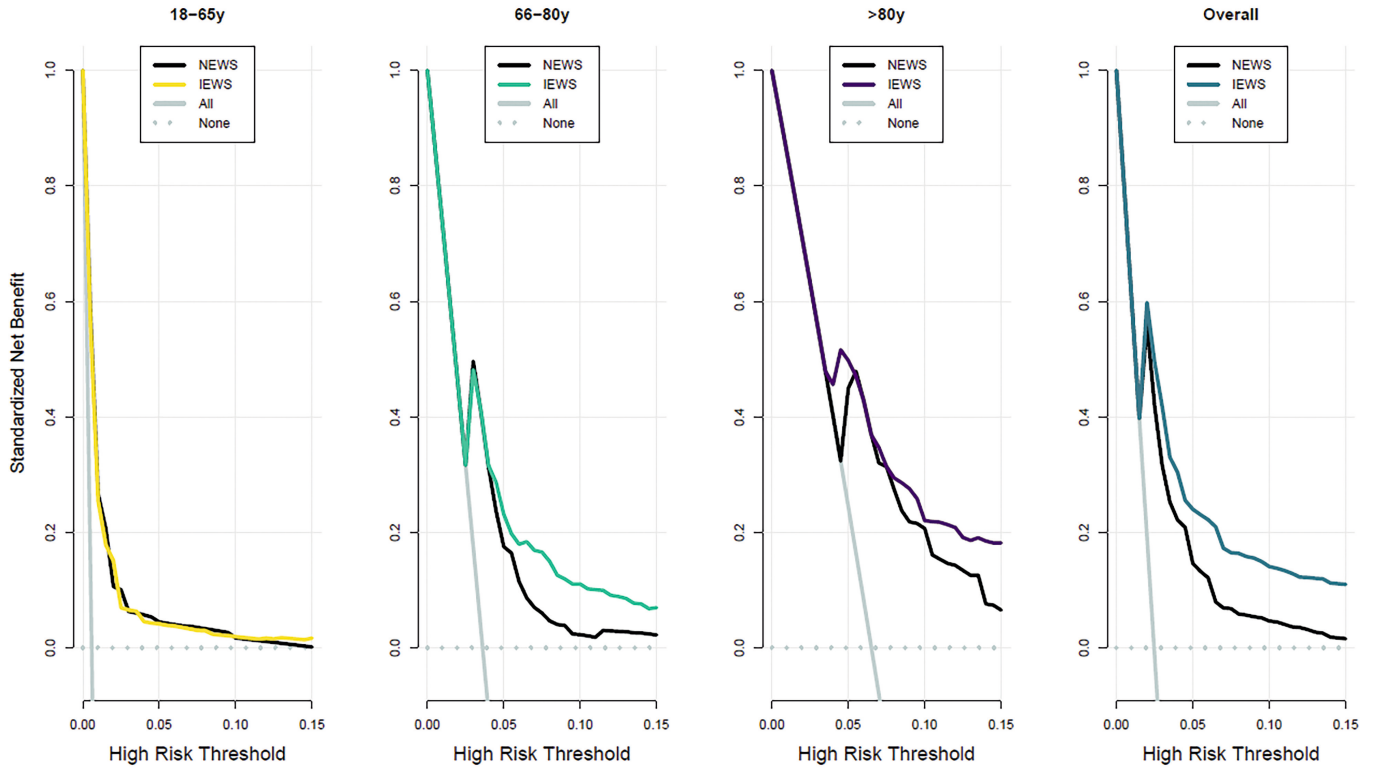
The figure below is a calibration plot for NEWS compared to IEWS stratified by hospital location.



Hospital 1 (yellow), Hospital 2 (green), hospital 3 (purple)

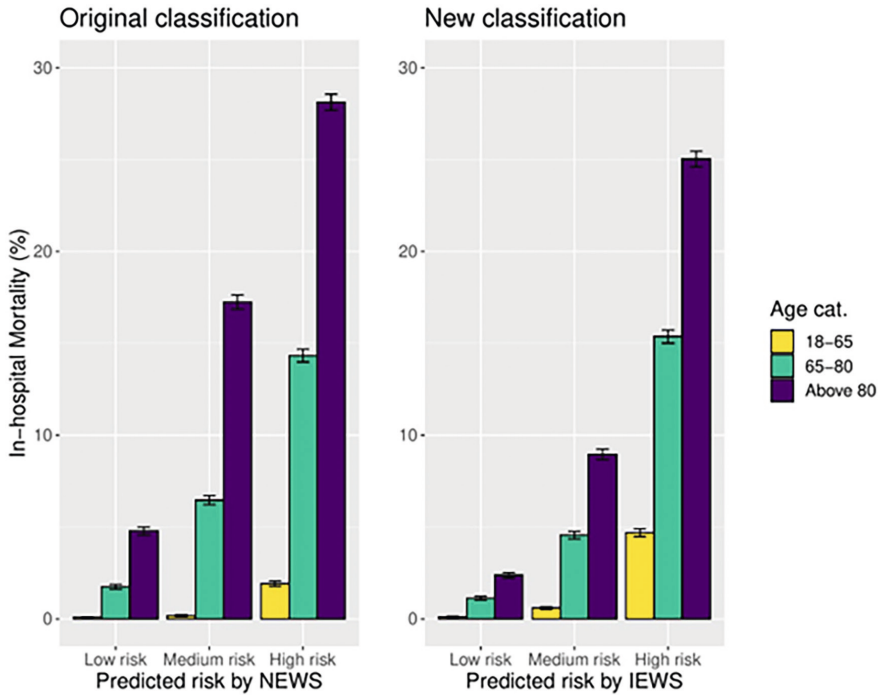
SUPPLEMENTARY FILE 12

Decision Curve Analysis in the DMC



SUPPLEMENTARY FILE 13

Reclassification figure and table for the development cohort.
Because no clear risk thresholds exist, the baseline risk was calculated for each age category by the average mortality risk for patients with a NEWS<4. Patients were considered as lows risk if mortality was lower than 2*baseline risk, medium risk if mortality was between 2*baseline risk and 3*baseline risk, and high risk if mortality was more than 3*baseline risk.
18-65y Low risk <0.68%, medium risk 0.68-1.02% and high risk >1.02%
66-80y Low risk <2.94%, medium risk 2.94-4.41% and high risk >4.41%
>80y Low risk <5.96%, medium risk 5.96-8.94% and high risk >8.94%



	Classification by NEWS			Classification by IEWS		
	Low risk (N45218)	**Medium Risk (N=16657)**	**High risk (N=34178)**	**Low risk (N=66700)**	**Medium risk (N=10588)**	**High risk (N=18765)**
Age category						
18-65y (alive)	9359 (20.7%)	13735 (82.5%)	27991 (81.9%)	34486 (51.7%)	6355 (60.0%)	10244 (54.6%)
18-65y (died)	8 (0.0%)	23 (0.1%)	546 (1.6%)	36 (0.1%)	38 (0.4%)	503 (2.7%)
66-80y (alive)	22631 (50.0%)	2115 (12.7%)	3864 (11.3%)	21379 (32.1%)	2704 (25.5%)	4527 (24.1%)
66-80y (died)	404 (0.9%)	146 (0.9%)	646 (1.9%)	245 (0.4%)	129 (1.2%)	822 (4.4%)
>80y (alive)	12203 (27.0%)	528 (3.2%)	813 (2.4%)	10303 (15.4%)	1240 (11.7%)	2001 (10.7%)
>80y (died)	613 (1.4%)	110 (0.7%)	318 (0.9%)	251 (0.4%	122 (1.2%)	668 (3.6%)

SUPPLEMENTARY FILE 14

Multivariable logistic regression for missing glasgow coma scale**:** A multivariable logistic regression model demonstrated that missing glasgow coma scale is associated with physiological variables and outcome (Chi-square p<0.01), and thus GCS is not missing completely at random or missing not at ramdom and is suitable for multiple imputation. Similar Chi-square p-values were found for the other physiological variables.
Logistic Regression Model
lrm(formula = is.na(gcs) ~ ihm + sbp + saturation + sex + triage +
age, data = NEED2)

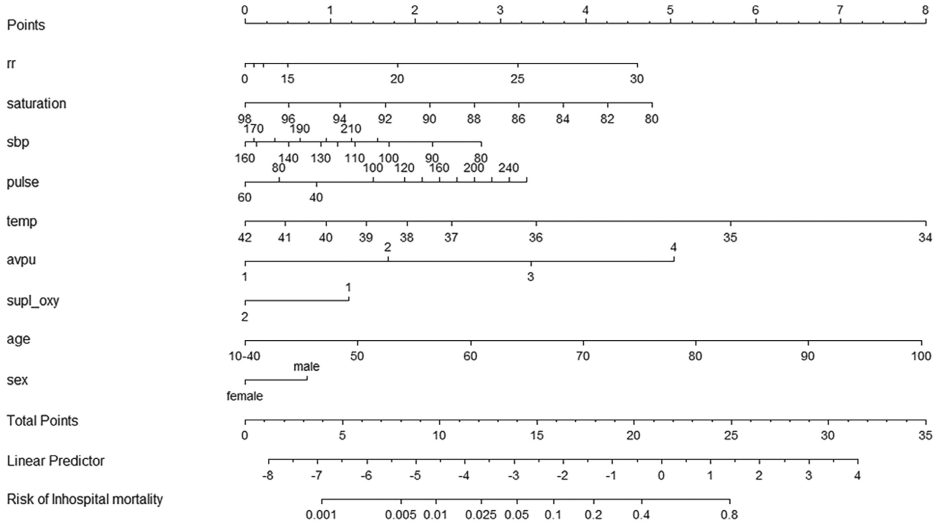
Ratio Test
Obs 82306
FALSE 9290
TRUE 73016 Pr(> chi2) <0.0001

Coef	S.E.		Wald Z Pr(>|Z|)
Intercept	2.7829 0.3175	8.77	<0.0001
ihm	-0.2334 0.0630	-3.70	0.0002
sbp	-0.0095 0.0004	-26.49	<0.0001
saturation	0.0032 0.0012	0.37	0.7112
sex=female	0.0375 0.0225	1.67	0.0954
triage=very urgent	0.4236 0.0394	10.76	<0.0001
triage=urgent	0.9677 0.0397	24.37	<0.0001
triage=non-urgent	1.6921 0.0495	34.17	<0.0001
age	-0.0058 0.0007	-8.78	<0.0001

Frequencies of Missing Values Due to Each Variable
is.na(gcs) ihm sbp saturation sex triage age
0 2154 59317 57074 2 6181 0

SUPPLEMENTARY FILE 15

Nomogram for vital signs used as restricted cubic splines. Extremes for each vital sign were excluded.





PART III

BIAS IN THE USE OF RISK TOOLS TO COMPARE QUALITY OF CARE


EMERGENCY

9

THE EFFECT OF TREATMENT AND CLINICAL COURSE DURING EMERGENCY DEPARTMENT STAY ON SEVERITY SCORING AND PREDICTED MORTALITY RISK IN INTENSIVE CARE PATIENTS.

Bart GJ Candel
Wouter Raven
Heleen Lameijer
Wendy AMH Thijssen
Fabian Termorshuizen
Christiaan Boerma
Nicolette F. de Keizer
Evert de Jonge
Bas de Groot

Published in Critical Care – 19 April 2022 Volume 26, Issue 1, page 1-9.

ABSTRACT

Background Treatment and the clinical course during Emergency Department (ED) stay before Intensive Care Unit (ICU) admission may affect predicted mortality risk calculated by the Acute Physiology and Chronic Health Evaluation (APACHE)-IV, causing lead-time bias. As a result, comparing standardized mortality ratios (SMRs) among hospitals may be difficult if they differ in the location where initial stabilization takes place. The aim of this study was to assess to what extent predicted mortality risk would be affected if the APACHE-IV score was recalculated with the initial physiological variables from the ED. Secondly, to evaluate whether ED Length of Stay (LOS) was associated with a change (delta) in these APACHE-IV scores.

Methods An observational multicenter cohort study including ICU patients admitted from the ED. Data from two Dutch quality registries were linked: the Netherlands Emergency department Evaluation Database (NEED) and the National Intensive Care Evaluation (NICE) registry. The ICU APACHE-IV, predicted mortality, and SMR based on data of the first 24 hours of ICU admission were compared with a ED APACHE-IV model, using the most deviating physiological variables from the ED or ICU.

Results A total of 1398 patients were included. The predicted mortality from the ICU APACHE-IV (median 0.10; IQR 0.03-0.30) was significantly lower compared to the ED APACHE-IV model (median 0.13; 0.04-0.36; p<0.01). The SMR changed from 0.63 (95%CI 0.54-0.72) to 0.55 (95%CI 0.47-0.63) based on ED APACHE-IV. Predicted mortality risk changed more than 5% in 321 (23.2%) patients by using the ED APACHE-IV. ED LOS >3.9hours was associated with a slight increase in delta APACHE-IV of 1.6 (95% CI 0.4-2.8) compared to ED LOS <1.7hours.

Conclusion Predicted mortality risks and SMRs calculated by the APACHE IV scores are not directly comparable in patients admitted from the ED if hospitals differ in their policy to stabilize patients in the ED before ICU admission. Future research should focus on developing models to adjust for these differences.

BACKGROUND

The Acute Physiology and Chronic Health Evaluation (APACHE)-IV is a scoring system used for risk stratification of Intensive Care Unit (ICU)-patients.[32, 196] The APACHE-IV is also used for case-mix correction when comparing outcomes of patients in national and international research or quality improvement projects.

The acute physiology score (APS), part of the APACHE-IV model, uses the most deviated vital signs and laboratory results measured in the first 24 hours after ICU admission.[31, 32] However, physiological parameters may improve or deteriorate in the hours before ICU admission,[33-36], e.g., during initial treatment in the emergency department (ED) or because of deterioration in the clinical course. These data measured prior to ICU admission are not included in the original APACHE-IV scores.[31, 32] Consequently, the calculated severity of illness and predicted mortality risk may differ when for a given patient the same initial treatment is given before or after ICU admission or if the ED stay is shorter resulting in a different clinical course, a phenomenon called lead-time bias in literature.[37, 38] For appropriate risk stratification and case-mix correction, it is essential to understand whether there is an effect of ED treatment on the APACHE-IV scores.

The present study assessed if, and to what extent, the APACHE-IV score and predicted mortality risk change if the APACHE-IV is recalculated, including data from both the pre-ICU ED period and the first 24 hours of ICU treatment. In this way we evaluated if earlier findings from a very small sample could be confirmed.[38] Our second aim was to assess whether ED treatments (e.g., the amount of fluid administered) and clinical course during ED stay were associated with changes in the APACHE-IV score, predicted mortality risk, and standardized mortality ratio (SMR).

METHODS

Study design and setting

This was an observational multicenter cohort study using the Netherlands Emergency department Evaluation Database ((NEED), www.stichting-need.nl) and the National Intensive Care Evaluation (NICE) registry.[197] Data were available from two urban teaching hospitals (Medical Center Leeuwarden, 14 October 2017 – 31 July 2020 and Catharina Hospital Eindhoven, 01 January 2019 – 01 July 2020). In both hospitals, patients were treated and stabilized in the ED before ICU admission by emergency physicians with or without consultation of an intensivist. If necessary, mechanical ventilation was

started in the ED. Patients were almost never intubated prehospital because hospitals can be reached within 15minutes by ambulance in both regions in the Netherlands. The medical ethical committee of Máxima MC reviewed the research proposal and concluded that the anonymized data were not subject to the Dutch Research on Humans Subjects Act (in Dutch "WMO") and waived the need for informed consent (registration number N20.117).

Patient selection

All ICU patients who were directly admitted from the ED in the mentioned study period were included. Patients transferred to other hospitals and patients without any registered vital signs or laboratory results (9 patients) in the NEED were excluded.

Data collection

Data from the NICE

The NICE is the national quality registry founded in 1996 in which all ICUs in the Netherlands have participated since 2016.[198] Data available from NICE were: Age, sex, date of admission, ICU length of stay, ICU mortality, hospital mortality, and all data to calculate the APACHE-IV score.[32, 54] Supplementary file 1 shows which variables were extracted from the NICE and from the NEED to calculate the APACHE-IV score. Only the lowest and the highest values of vital signs and laboratory tests in the first 24hours of ICU admission were registered in the NICE. The NICE registry uses strict definitions and specifications of the data collected, described in a data dictionary (https://www.stichting-nice.nl/dd/). The NICE provides participants a mandatory training and performs automated checks on data entry and regular on-site quality audits.[198, 199]

Data from the NEED

The NEED is the quality registry for EDs in the Netherlands (www.stichting-need.nl), founded in 2016. For the present study, data were available from two hospitals. Only the first set of vital signs and laboratory results were registered in the NEED, measured before ED treatment, described in detail previously.[82] Except for variables to calculate the APACHE-IV score (Supplementary file 1), patient characteristics (e.g., triage category according to the Dutch Triage System or Manchester Triage Standard, presenting complaints, diagnostics (Y/N), fluid administration), and ED- length of stay (LOS) were extracted.

Linking of NEED and NICE databases
Both NEED and NICE have unique identifying patient numbers that are not identical. In the NICE, patients were selected who were admitted from the ED. In the NEED, patients were selected who were admitted to a high dependency care unit. If the following variables were identical between both databases, patient variables were linked: Date of ED admission (in the NEED) and the date of ICU admission (in the NICE), hospital location, sex, and age of the patient. In addition, for records in NICE with ED as resource but without linkage to a NEED record, the date of ED admission could be one day earlier than the date of ICU admission. These patients were linked by changing the data of ED presentation to one day earlier.

Outcome measures
Outcome measures were the original ICU APACHE-IV score (Formally the APACHE III score used in the APACHE IV model [31, 32]) and predicted mortality using the most deviated physiological variables from the first 24 hours after ICU admission, and the ED APACHE-IV score using the most deviated values from ED admission until 24 hours after ICU admission. The SMR is the ratio between the observed number of deaths and the expected deaths in the ICU using the original (ICU model) or modified (ED model) predicted mortality risks. The delta APACHE-IV was calculated as the ED APACHE-IV score – ICU APACHE-IV score.

Data and data analyses

Descriptive statistics
Data are presented as mean with standard deviation (SD) if normally distributed and as median with interquartile range (IQR) if skewed. Categorical data are presented as frequencies (%).

Main statistical analyses
In our study, the ICU APACHE-IV score was compared with a recalculated APACHE IV score, called the ED APACHE-IV. According to the original model, the ICU APACHE-IV score was calculated with the most deviating vital signs or blood tests in the first 24hours of ICU admission.[31] Points were assigned for the value that was furthest from a reference value. For example, for heart rate the reference value was 75bpm. Reference values for all other variables are shown in supplementary file 2. According to the APACHE IV definitions,[31] more deviating vital signs or laboratory results do not necessarily result in a higher score (see supplementary file 2). Based on rules of the original APACHE-IV model, missing physiological variables were considered normal. The ED APACHE-IV included the most deviating vital

signs and blood tests from ED admission until 24 hours after ICU admission. The data available from the NEED registry to calculate the ED APACHE-IV are described in Supplementary file 1. The acute physiological variables were recalculated, all other variables in the ED APACHE-IV remained identical to the ICU APACHE-IV (e.g., the Chronic Health Conditions). Urine output and mechanical ventilation (Y/N) were not registered in the NEED. Patients were considered not to be intubated at ED arrival. In the ED, venous blood gas is often obtained instead of an arterial blood gas. If no arterial blood gas was available, the pH from the venous blood gas analysis was used with a correction of 0.03, and the pCO2 was used with a correction of -4.8mmHg.[200] PaO2 is not well correlated between venous and arterial blood gas and was therefore considered missing if not available as arterial value. Predicted mortality risks were calculated for the ED APACHE-IV using the Beta's of the ICU APACHE-IV logistic model. The APACHE-IV scores, Acute Physiology Score (APS), and predicted mortality risks were compared between the ICU and ED model with a Wilcoxon Signed Rank test. In addition, we assessed the discriminative performance of both the ICU and ED APACHE-IV model using a receiver operator characteristic (ROC) curve with area under the curve (AUC) analysis and in-hospital mortality as outcome. Calibration plots for both models were presented.

For our second aim, delta APACHE-IV was calculated (e.g., ED APACHE-IV – ICU APACHE-IV). Delta APACHE-IV was normally distributed, and therefore, univariable and multivariable linear regression analyses were performed to assess the association between ED LOS, fluid administration (0ml, 0-500ml, >500ml) as independent variables, and a delta APACHE-IV as dependent variable. ED-LOS was divided into quartiles (0-1.7h, 1.7-2.7h, 2.7-3.9h, >3.9h). Because of a strong correlation between ED LOS and fluid administration, separate models were used. The following potential confounders were included in both multivariable models: Age, gender, hospital, and the ED APACHE-IV score. Although age is one of the variables included in the APACHE-IV, age may still be an independent predictor for delta APACHE-IV. Fluid administration and ED LOS were included as dummy variables to overcome non-linear associations. The unstandardized coefficients (Beta's) were presented with 95% Confidence Intervals (95%-CI).

To study whether the ED APACHE-IV was significantly different per quartile ED-LOS, a Kruskal Wallis Test was used.

A P-value <0.05 was considered statistically significant. SPSS version 25.0 was used for all data analyses.

RESULTS

Patient characteristics

During the study period, 1730 patients were admitted from the ED to the ICU. In total, 1398 (80.8%) patients were included after linking both databases (see figure 1). The median age was 64 years (50-74years), a total of 1046 (60.5%) were male patients. Patient characteristics, including all APS variables, are described in table 1. Characteristics of patients who could not be linked between both databases (N=323) are comparable with the included patients and are described in supplementary file 3. The median laboratory values in the ED were comparable with the lowest and highest values from the ICU. Supplementary file 4 describes other patient characteristics, such as the triage category in the ED, and the most common presenting complaints in the ED and reasons for ICU admissions.

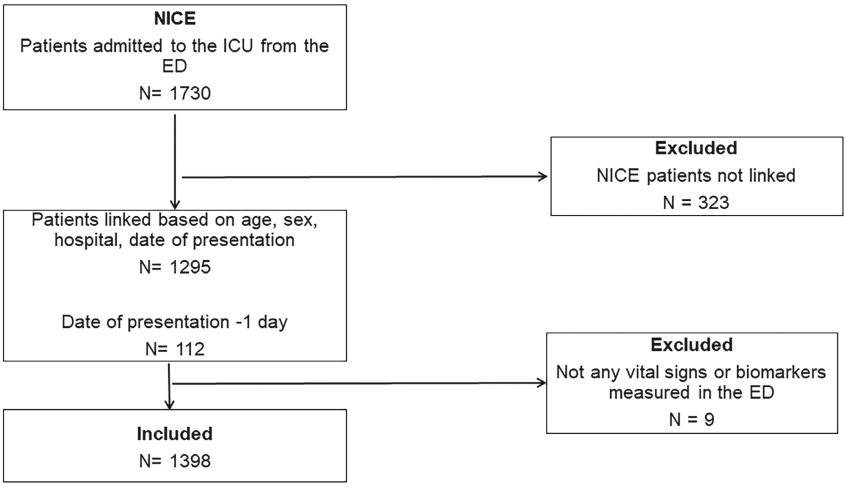


Figure 1. Patient flow diagram throughout the study. Patients from the National Intensive Care Evaluation (NICE) registry were linked with patients from the Netherlands Emergency department Evaluation Database (NEED).

Table 1 Patient characteristics

Characteristics (N=1398)	**Emergency Department**	**Intensive Care Unit**	**Intensive Care Unit**
Age, years, median (IQR)	64 (50-74)	-	-
Sex, male, N (%)	1046 (60.5)	-	-
Vital signs, median (IQR) {missing}	**Initial values**	**Lowest values <24h**	**Highest values <24h**
MAP, mmHg	97 (80-114) {159}	62 (53-71) {6}	100 (88-115) {4}
HR, bpm	95 (80-114) {156}	71 (59-83) {3}	103 (88-119) {1}
SpO2, %	97 (94-100) {170}	-	-
RR, /min.	20 (16-25) {223}	13 (10-16) {8}	25 (22-29) {10}
Temperature, ºC	36.7 (36.1-37.3) {429}	36.4 (35.8-36.9) {12}	37.3 (36.8-37.9) {18}
GCS	15 (11-15) {962}	15 (10-15)	
Urine, 24hours, L		1.3 (0.9-2.0)	
Biomarkers, median (IQR) {missing}			
Creatinine, µmol/L	88 (71-120) {77}	81 (64-116) {156}	87 (68-128) {164}
Urea, mmol/L	6.5 (4.7-9.4) {76}	7.2 (5.0-10.5) {167}	
Hematocrit, L/L	0.42 (0.37-0.46) {52}	0.36 (0.31-0.40 {115}	0.39 (0.34-0.43) {122}
Leukocytes, x10^9/L	11.6 (8.2-16.2) {79}	10.9 (7.8-15.0) {225}	12.3 (8.8-17.4) {229}
Sodium, mmol/L	139 (136-142) {53}	137 (134-140) {118}	140 (137-143) {125}
Albumin, g/L	42 (37-45) {828}	29 (24-34) {583}	30 (24-34) {585}
Glucose, mmol/L	7.8 (6.3-11.3) {824}	6.0 (5.2-7.2) {116}	8.8 (7.0-11.6) {124}
Bilirubin, µmol/L	10 (6-15) {431}	8.9 (5.5-14) {547}	

Table 1 Patient characteristics (continued)

Characteristics (N=1398)	Emergency Department	Intensive Care Unit	Intensive Care Unit
Blood gas, median (IQR) {missing}			
a-PO2, mmHg	86 (60-145) {687}	82 (70-102) {379}	
a/v-PCO2, mmHg	40 (35-50) {346}	39 (33-45) {379}	
a/v-pH	7.36 (7.27-7.43) {344}	7.40 (7.30-7.41) {375}	
Mechanical ventilation, N (%)	-	-	611 (43.7)
FiO2 (%), median (IQR)	-	-	35 (25-45) {224}

ED: Emergency Department, ICU: Intensive Care Unit, IQR: Interquartile Range, N: number, MAP: mean arterial pressure, mmHg: millimeter mercury, HR: heart rate, SpO2: peripheral oxygen saturation, RR: respiratory rate, ^{o}C: degrees Celsius, GCS: Glasgow Coma Scale, L: Liter, a-PO2: arterial partial pressure of oxygen, a/v-PCO2: arterial or venous partial pressure of carbon dioxide, a/v-pH: arterial or venous acid base, FiO2: Fraction of inspired oxygen.
The initial values from the ED are presented before ED treatment. From the ICU, both the lowest and highest values are given from the first 24hours after admission, if available.
If arterial blood gas analysis was not available in the ED, venous PCO2 was used with a correction of -4.8mmHg, and pH was used with a correction of 0.03.

The ED APACHE-IV score

Including the most deviating vital signs and blood tests from ED admission until 24 hours after ICU admission, the median ED APACHE-IV score (63; IQR 47-90) was calculated and differed significantly from the median ICU APACHE-IV score (56; IQR 39-80) (p-value <0.01). The median predicted mortality for the total population was higher for the ED APACHE-IV system, 0.13 (IQR 0.04-0.36) versus 0.10 (IQR 0.03-0.30) (see table 2). Observed in-hospital mortality was 13.4% (N=188). The SMR decreased from 0.63 (95% CI 0.54-0.72) for the ICU APACHE-IV model to 0.55 (95% CI 0.47-0.63) for the ED APACHE-IV model, a relative difference of 12.7%.

Table 2 The ICU and ED APACHE-IV model

Risk scores	ICU APACHE-IV model	ED APACHE-IV model	P-value
APACHE-IV score, median (IQR)	56 (39-80)	63 (47-90)	<0.01
APS, median (IQR)	44 (30-68)	52 (35-76)	<0.01
APACHE-IV Predicted mortality, median (IQR)	0.10 (0.03-0.30)	0.13 (0.04-0.36)	<0.01
SMR (95% CI)	0.63 (0.54-0.72)	0.55 (0.57-0.63)	

ICU: Intensive Care Unit, ED: Emergency Department, APACHE-IV: Acute Physiology and Chronic Health Evaluation (4th edition), APS: Acute Physiology Score, SMR: Standardized mortality Ratio (observed mortality / predicted mortality). IQR: interquartile range, 95% CI: 95 percent Confidence Interval.
The ED APACHE-IV was calculated by using the most deviating vital signs and laboratory results from the Emergency Department or the first 24hours of Intensive Care Unit admission. All other variables in the APACHE-IV model remained similar.

Both APACHE-IV models had an identical AUROC of 0.91 (95% CI 0.89-0.93) and 0.91 (95% CI 0.89-0.93). Calibration plots for both models were comparable (supplementary file 5). The ED APACHE-IV score increased in 1075 (77.2%) patients, decreased in 33 (2.4%) patients, and remained unaltered in 284 (20.4%) patients. Predicted mortality increased ≥5.0% in 320 (23.1%) patients and decreased ≥5.0% in 1 (0.1%) patient (see Supplementary file 6). Higher ICU APACHE-IV scores had larger changes in predicted mortality risks if modified with ED variables.

ED Length of stay and delta APACHE-IV score

To study whether the ED LOS was associated with a delta APACHE-IV score (= ED APACHE-IV – ICU APACHE-IV), the ED APACHE-IV score and the delta APACHE-IV are presented per ED LOS quartile (figure 2). As assessed with a multivariable linear model, more fluid administration in the ED

was associated with a significant adjusted increase in delta APACHE-IV (p-value=0.04) (see table 3). Compared to those with an ED LOS < 1.7hours, among patients with ED LOS >3.9hours the delta-APACHE was on average 1.6 points (95% CI 0.4-2.8) higher.
The median APACHE-IV score differed per quartile ED LOS (p<0.01) calculated with the Kruskal Wallis Test.

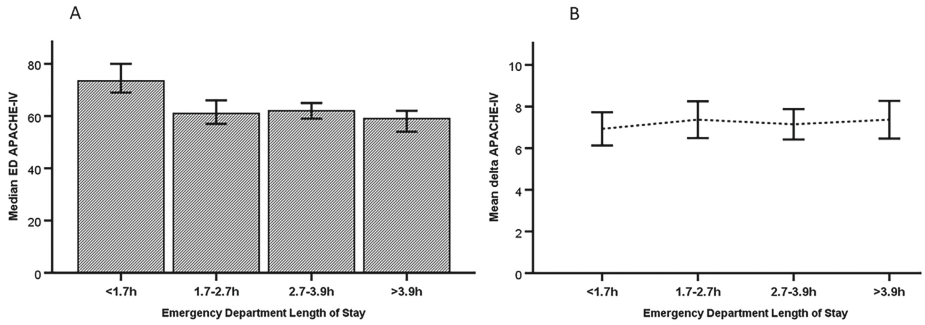


Figure 2 In panel A the median Emergency Department Acute Physiology and Chronic Health Evaluation (ED APACHE-IV) score is presented per quartile Emergency Department Length of Stay (ED-LOS), with 95% Confidence intervals. The ED APACHE-IV score uses the most deviated physiological variables from ED admission until 24 hours after ICU admission, which differs from the ICU APACHE-IV score which only contains the most deviated physiological variables from the first 24hours of ICU admission. Panel B shows the mean delta APACHE-IV which is calculated as follows: ED APACHE-IV score – ICU APACHE-IV score.

Table 3 Crude and adjusted associations between fluid administration, Emergency Department-Length of Stay and delta APACHE-IV

Independent variables Beta (95% CI)	Crude	P-value	Adjusted Beta (95% CI)	P-value
No fluid	-	-	-	-
0-500ml fluid	0.9 (-1.0-1.2)	0.9	0.6 (-0.6-1.8)	0.33
>500ml fluid	1.9 (1.0-2.8)	<0.01	1.1 (0.0-2.1)	0.04
ED-lOS <1.7h	-	-	-	-
ED-LOS 1.7-2.7h	-0.4 (-1.3- 0.6)	0.45	0.8 (-0.3- 2.0)	0.15
ED-LOS 1.7-3.9h	0.2 (-0.7-1.2)	0.64	0.9 (-0.2-2.1)	0.20
ED-LOS >3.9h	-0.1 (-1.0-0.9)	0.87	1.6 (0.4-2.8)	0.01

ED-LOS: Emergency Department Length of Stay, APACHE-IV: Acute Physiology and Chronic Health Evaluation (4th edition), 95% CI, 95 percent confidence intervals.
The association between fluid administration, ED-LOS and Delta APACHE-IV score (ED APACHE-IV – ICU APACHE-IV) was assessed. Crude Beta's are presented, and adjusted Beta's. Because of multicollinearity between ED LOS and fluid administration, separate models were used for both variables. Both models were adjusted for age, sex, hospital, and the modified APACHE-IV score.

DISCUSSION

The present study shows that in ICU patients admitted from the ED, the average calculated severity of illness and predicted mortality risks are substantially higher, and the SMR is lower if not only data from the first 24 hours of ICU stay are used to calculate the APACHE-IV score, but also the data from the ED. In our population, the mean increase in APACHE IV score was 13%, with a mean increase in predicted mortality risk of 30% and a decrease in SMR of 13%.

Our results are in accordance with an earlier study of 76 ICU patients admitted from the ED and operating rooms.[38] In this study, higher APACHE II, APACHE III, and SAPS II scores were found if calculated with data from 6 hours before ICU referral to 24 hours after ICU admission compared with the standard period from ICU admission to 24 hours after ICU admission. Consequently, including pre-ICU data leads to higher severity of illness scores. This can be explained by the fact that the period to choose the most deviating physiological variables becomes longer.

It may be argued that this influence of pre-ICU data on the assessment of severity of illness and predicted mortality has no major consequences,

because the 24h time period for data collection is clearly defined and similar for ICUs all over the world. Thus, APACHE scores and mortality predictions would be well comparable for describing case-mix and outcomes among ICUs (benchmarking). However, the true importance of our findings lies in the fact that major differences exist in the location where initial stabilization of patients takes place. For example, patients with septic shock are stabilized in the ED in some hospitals,[201] whereas they would be transferred immediately to the ICU in other hospitals.[37, 202, 203] Especially in the early phase of stabilization, physiological parameters may be very disturbed leading to high APACHE scores. Because the APACHE prognostic scores are calculated from data exclusively from the first 24 hours after ICU admission, this would contribute to higher APACHE scores and predicted mortality if stabilization took place in the ICU, but not if performed in the ED. Calculated severity of illness with the APACHE-IV score and predicted mortality risk will be lower in hospitals where patients are routinely stabilized in the ED before transfer to the ICU, compared to hospitals where patients are immediately transferred from the ED to the ICU and stabilized there.[203] As a result, the higher calculated SMRs may be falsely interpreted as a difference in the quality of care, while they can be fully explained by the fact that only data from the ICU are used for scoring. In our population, we found a substantial difference in SMR of 0.63 versus 0.55 depending on using or not using the ED data. In ICU literature this occurrence is referred to as lead-time bias.[37, 38]

Our findings imply that whenever the quality of care among ICUs is compared using the SMR, a correction should be applied, or a margin of error of approximately 13% should be used if hospitals differ in their policy to stabilize patients. Our study was performed only in patients admitted to the ICU from the ED. However, in patients admitted from the operating rooms, wards, or in patients transferred from other hospitals comparable differences in measured severity of illness and SMRs may be found if policies in stabilizing patients differ.[37]

In the ICU APACHE-IV model, the calculated predicted mortality risk is already adjusted for the source of admission, e.g., from the ED. However, this adjustment doesn't correct for the influence of initial stabilization in either the emergency department or the ICU. In both situations, the patient is admitted from the ED with identical adjustment for source of admission. Future studies should investigate whether the APACHE-IV model could be improved to account for differences in hospital policy to stabilize patients.

We show that patients with higher ICU APACHE-IV scores had a larger absolute difference in predicted mortality compared to patients with a lower

ICU APACHE-IV score. This may be explained by more intense resuscitative treatments in the ED. Unfortunately, the data available about therapies started in the ED was limited to only fluid therapy.

Small increases in delta APACHE-IV were found with more fluid administration in the ED and with an ED LOS of >3.9hours. We can only speculate why a longer ED LOS led to a larger difference in APACHE-IV. Likely, patients with a prolonged ED LOS got more treatments in the ED, such as fluid administration, but also unregistered resuscitative therapies, which caused bigger changes in physiological signs than in patients who had a short ED LOS. However, we expected to find larger differences between the ICU APACHE-IV and ED APACHE-IV with more intensive ED treatment and a longer ED LOS. Nonetheless, this study analyzed a selective group of patients from the ED who may not have responded sufficiently to therapy and therefore may have been admitted to the ICU. This selection of ED patients may explain why we found only small increases in delta APACHE-IV in patients with a prolonged ED LOS.

Despite several strengths like the sample size and the multicenter design, our study also has some limitations. First, in approximately 18% of patients admitted from the ED to the ICU, no ED data were available. Patients could not be linked between databases for various reasons: i.e., transfers to other hospitals, patients who first went for surgery and not directly to the ICU, patients who bypassed the ED and went for coronary intervention, and possible registration problems. We cannot exclude that this may lead to some selection bias. Nonetheless, patient characteristics and mortality of patients who could not be linked between both databases were comparable with characteristics of included patients. Also, in our ED database, only initial values from ED admission were present. It may well be that even more deviating values could have been measured during ED treatment. Therefore, our findings likely underestimate the actual so-called lead-time bias in our patients. Furthermore, in our study, we did not include patients admitted to the ICU from an ED in another hospital. In those cases, the period between ED admission and ICU admission is prolonged, which potentially leads to more considerable differences in calculated APACHE IV scores and thus more lead-time bias.[37]

CONCLUSIONS

In summary, including the initial vital signs and laboratory results from the ED in the APACHE-IV score changed the predicted mortality risk and SMR, highlighting the influence of treatment given at the ED prior to IC admission.

In addition, a longer ED LOS was associated with an increase in delta APACHE IV score, leading to a spuriously high SMR when this is not considered. Our findings are essential when comparing hospitals regarding quality of care and case-mix adjusted outcomes. APACHE IV scores and SMRs are not directly comparable in patients admitted from the ED if hospitals differ in their policy to stabilize patients in the ED before ICU admission. Future research should focus on refining models to adjust for these differences.

SUPPLEMENTARY FILE 1

Table to calculate the Acute Physiology Score (APS) as part of the Acute Physiology and Chronic Health Evaluation (APACHE)-IV.

Variables	Registered in NICE	Registered in NEED
Age	Yes	Yes
Acute Physiology Score (APS)		
- Heart rate	Yes	Yes
- Mean arterial pressure	Yes	Yes
- Temperature	Yes	Yes
- Respiratory Rate	Yes	Yes
- PaO2 (non-intubated patients, or intubated with FiO2<0.5)	Yes	Yes patients were considered not intubated at ED arrival
- P(A-a)-O2 for intubated patients with FiO2 >0.5	Yes	No
- PCO2	Yes	Yes If arterial PCO2 was not available, venous PCO2 was used with a correction of -4.8mmHg
- Hematocrit	Yes	Yes
- White blood cell count	Yes	Yes
- Creatinine	Yes	Yes
- Urine output	Yes	No
- Blood urea nitrogen	Yes	Yes
- Sodium	Yes	Yes

Table to calculate the Acute Physiology Score (APS) as part of the Acute Physiology and Chronic Health Evaluation (APACHE)-IV. (continued)

Variables	Registered in NICE	Registered in NEED
- Albumin	Yes	Yes
- Glucose	Yes	Yes
- Acid-base	Yes	Yes If arterial pH was not available, venous pH was used with a correction of 0.03
- Glasgow Coma Score (GCS)	Yes	Yes
Chronic Health variables		
- AIDS	Yes	No
- Cirrhosis	Yes	No
- Hepatic failure	Yes	No
- Immunosuppression	Yes	No
- Lymphoma, leukemia, or myeloma	Yes	No
- Metastatic tumor	Yes	No
ICU admission diagnosis	Yes	No
ICU admission source (e.g., the ED)	Yes	No
Emergency surgery (Y/N)	Yes	No
Thrombolytic therapy for patients with myocardial infarction	Yes	No
Mechanical ventilation	Yes	No

SUPPLEMENTARY FILE 2

Calculation of the Acute Physiology Score: In the left column the variables of the Acute Physiology Score as part of the APACHE-IV are given with their reference value. Points are assigned for values that are furthest away from this reference value.

Pulse (beats/min)	<=39	8	Pulse points
Select heart rate furthest from 75	40-49	5	
	50-99	0	
	100-109	1	
	110-119	5	
	120-139	7	
	140-154	13	
	>=155	17	
Mean Blood Pressure (MAP)	<=39	23	MAP points
Select MAP furthest from 90	40-59	15	
	60-69	7	
	70-79	6	
	80-99	0	
	100-119	4	
	120-129	7	
	130-139	9	
	>=140	10	
Temperature (degrees Centigrade)	<=32.9	20	Temperature points
Select core temperature furthest from 38.	33-33.4	16	
Add 1 degree Centigrade to axillary temps	33.5-33.9	13	

prior to selecting worst value	34-34.9	8	
	35-35.9	2	
	36-39.9	0	
	>=40	4	
Respiratory Rate (breaths/min)*	<=5	17	Respiratory rate points
Select respiratory rate furthest from 19	6-11	8	
**for patients who are ventilated*	12-13	7	
no points are given for respiratory	14-24	0	
rates of 6-12.	25-34	6	
	35-39	9	
	40-49	11	
	>=50	18	
PaO2 (mmHg)*	<=49	15	PaO2 points
**Use only for non-intubated patients or*	50-69	5	
intubated patients with FiO2 <0.5 (50%).	70-79	2	
	>=80	0	**OR**
OR			
A-aDO2*	<100	0	A-aDO2 points
**Only use A-aDO2 for intubated patients*	100-249	7	
with FiO2 >=0.5 (50%). Do not use PaO2	250-349	9	
weights for these patients.	350-499	11	
	>=500	14	

Hematocrit (%)	<=40.9	3	Hematocrit points
Select hematocrit furthest from 45.5.	41-49	0	
	>=50	3	
WBC (cu/mm)	<1.0	19	WBC points
Select WBC furthest from 11.5.	1.0-2.9	5	
	3.0-19.9	0	
	20-24.9	1	
	>=25	5	
Creatinine without ARF* (mg/dl)	<=0.4	3	Creatinine without ARF points
Select creatinine furthest from 1.0.	0.5-1.4	0	
	1.5-1.94	4	
OR	>=1.95	7	**OR**
Creatinine with ARF* (mg/dl)	0-1.4	0	Creatinine with ARF points
**Acute Renal Failure (ARF) is defined as creatinine >=1.5 mg/dl*	>=1.5	10	
as creatinine >=1.5 mg/dl			
and urine output <410 cc/day and no			
chronic dialysis			
Urine Output (cc/day)	<=399	15	Urine output points

Enter total for day	400-599	8	
	600-899	7	
	900-1499	5	
	1500-1999	4	
	2000-3999	0	
	>=4000	1	
BUN (mg/dl)	<=16.9	0	BUN points
Select highest BUN (furthest from 0)	17-19	2	
	20-39	7	
	40-79	11	
	>=80	12	
Sodium (mEq/L)	<=119	3	Sodium points
Select Sodium furthest from 145.5.	120-134	2	
	135-154	0	
	>=155	4	
Albumin (g/dl)	<=1.9	11	Albumin points
Select albumin furthest from 3.5.	2.0-2.4	6	
	2.5-4.4	0	
	>=4.5	4	
Bilirubin (mg/dl)	<=1.9	0	Bilirubin points
Select highest bilirubin (furthest from 0)	2.0-2.9	5	
	3.0-4.9	6	
	5.0-7.9	8	
	>=8.0	16	
Glucose (mg/dl)*	<=39	8	Glucose points

Select glucose furthest from 130	40-59	9			
**Glucose <=39 mg/dl is lower*	60-199	0			
weight than 40-59	200-349	3			
	>=350	5			
			Acid-Base points (see below)		
			GCS points (see below)		
			TOTAL APS Score	0	

**If a patient is anesthetized, under the influence of anesthesia, or totally paralyzed/sedated during the ENTIRE data collection period, attempt to obtain a GCS from the twelve-hour period prior to ICU admission when GCS was able to be assessed. If no assessable GCS is documented during that time period, zero points shhould be assigned. If unable to determine verbal score (due to intubation status or similar barriers), use clinical judgement and assign Glasgow Verbal Score according to the following scale:*

alert, oriented 5

confused 3

nonresponsive 1

If patient's eyes open spontaneously (4) or to	verbal				
painful/verbal stimulation (2,3)*, use scale:*	motor	oriented, converses (5)	confused conversation (4)	inappropriate words, incomp. sounds (3,2)	no response (1)

Note: shaded areas represent unlikely clinical combinations. Placing a patient in any of these cells should be done after careful confirmation of clinical findings.	obeys verbal commands (6)	0	3	10	15
	localizes pain (5)	3	8	13	15
	flexion withdrawal/ decorticate rigidity (4,3)	3	13	24	24
	decerebrate rigidity/no response (2,1)	3	13	29	29
If patient's eyes do not open spontaneously	verbal				
***or to painful/verbal stimulation (1)**, use scale:*	motor	oriented, converses (5)	confused conversation (4)	inappropriate words, incomp. sounds (3,2)	no response (1)
Note: shaded areas represent extremely unlikely clinical combinations, and should not be used. If these combinations are verified in a clinical setting, no prediction should be generated for the patient.	obeys verbal commands (6)				16
	localizes pain (5)				16

	flexion withdrawal/ decorticate rigidity (4,3)	24	33
	decerebrate rigidity/no response (1)	29	48

Acid-Base Abnormalities*

**Assign score based upon pH-pCO2 relationship*

	pCO2<25	25-<30	30-<35	35-<40	40-<45	45-<50	50-<55	55-<60	>=60
pH <7.15									
7.15-<7.2				12				4	
7.20-<7.25									
7.25-<7.30			6		3		2		
7.30-<7.35	9								
7.35-<7.40									
7.40-<7.45				0			1		
7.45-<7.50	5		0	2					
7.50-<7.55									
7.55-<7.60									
7.60-<7.65	0		3			12			
>=7.65									

SUPPLEMENTARY FILE 3

Patient characteristics of excluded patients who could not be linked with the Emergency Department Database (N=323)

Characteristics (N=323)	Intensive Care Unit	Intensive Care Unit
Age, years, median (IQR)	65 (53-74)	-
Sex, male, N (%)	221 (68.4)	-
In hospital mortality, N (%)	55 (17.0)	
Vital signs, median (IQR) {missing}	Lowest values <24h	Highest values <24h
MAP, mmHg	63 (53-72) {2}	99 (88-116) {3}
HR, bpm	71 (59-83) {3}	103 (88-119) {1}
RR, /min.	12 (10-15) {4}	24 (22-27) {5}
Temperature, ºC	36.3 (35.3-36.9) {2}	37.4 (36.7-37.9) {4}
GCS	15 (6-15)	
Urine, 24hours, L	1.2 (0.8-1.9) {28}	
Creatinine, µmol/L	82 (66-108) {26}	91 (71-120) {27}
Urea, mmol/L	7.0 (5.6-9.7) {30}	
Haematocrit, L/L	0.36 (0.32-0.41) {12}	0.41 (0.35-0.45) {13}
Leukocytes, x10^9/L	10.9 (8.1-14.4) {54}	12.1 (9.2-16.9) {55}
Sodium, mmol/L	137 (135-139) {17}	140 (138-142) {18}
Albumin, g/L	29 (25-34) {161}	30 (25-34) {161)
Glucose, mmol/L	6.2 (5.4-7.1) {18}	9.1 (7.0-11.8) {19}
Bilirubin, µmol/L	10 (7-16) {149}	
a-PO2, mmHg	88 (73-107) {103}	
a-PCO2, mmHg	38 (35-44) {103}	
a-pH	7.40 (7.30-7.40) {103}	
Mechanical ventilation, N (%)	-	170 (52.6)
FiO2 (%), median (IQR)	-	40 (30-50) {90}

ICU: Intensive Care Unit, IQR: Interquartile Range, N: number, MAP: mean arterial pressure, mmHg: millimeter mercury, HR: heart rate, SpO2: peripheral oxygen saturation, RR: respiratory rate, ºC: degrees Celsius, GCS: Glasgow Coma Scale, L: Liter, a-PO2: arterial partial pressure of oxygen, a/v-PCO2: arterial or venous partial pressure of carbon dioxide, a/v-pH: arterial acid base, FiO2: Fraction of inspired oxygen.
From the ICU, both the lowest and highest values are given from the first 24hours after admission, if available.

SUPPLEMENTARY FILE 4

Patient demographics including triage categories, top five presenting: complaints, top seven reasons for admission to the ED, diagnostics in the ED and fluid administration.

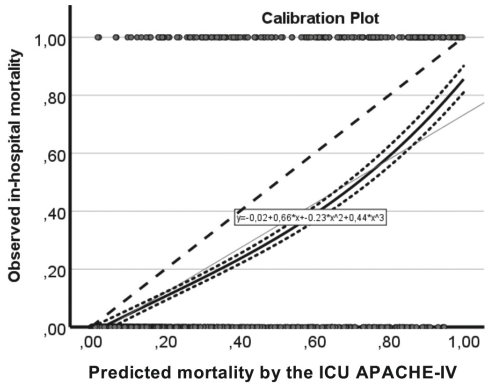
Demographics	**N = 1398**
Arrived by ambulance, N (%)	985 (81.4) {188}
Triage category in the ED, N (%)	{16}
Green/blue (non-urgent)	67 (5.0)
Yellow (less urgent)	340 (25.2)
Orange (urgent)	636 (47.2)
Red (Immediate)	305 (22.6)
Top five chief complaints in the ED, N (%)	{246}
Feeling unwell	222 (15.9)
Dyspnoea	216 (15.5)
Intoxication	154 (11.0)
Abdominal pain	130 (9.3)
Thoracic pain	84 (6.0)
Miscellaneous	592 (42.3)
Top seven reasons for ICU admission, N (%)	
Drug overdose	205 (14.7)
Sepsis	188 (13.4)
After cardiac arrest	136 (9.7)
Peripheral vascular surgery -surgical	87 (6.2)
Infection	82 (5.9)
Pulmonary embolism	80 (5.7)
Cardiovascular	75 (5.4)
Miscellaneous	545 (39.0)
Diagnostics in the ED, N (%)	
Blood cultures	335 (19.4)
ECG	992 (57.3)
Radiology	962 (68.8)

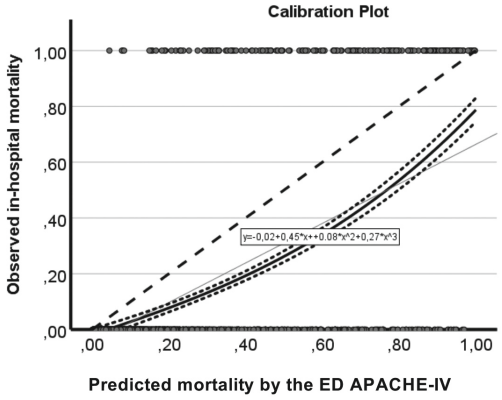
Fluid administration in the ED	
0 ml	1085 (62.7)
0-500ml	241 (13.9)
>500ml	404 (23.4)

ED: Emergency Department, ECG: Electrocardiogram
Numbers between {} represent missings.
Triage categories are based on the Manchester Triage System or the Dutch Triage Standard.

SUPPLEMENTARY FILE 5

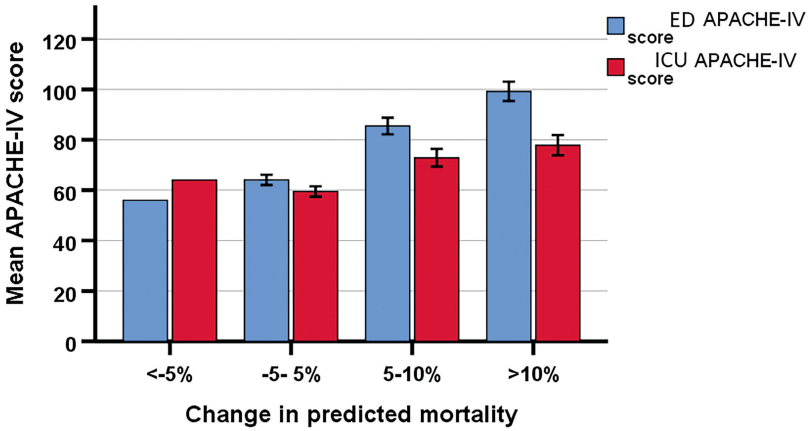
Calibration plots for both the ICU APACHE-IV score and the ED APACHE-IV score





SUPPLEMENTARY FILE 6

The change in predicted mortality is presented if the ED Acute Physiology and Chronic Health Evaluation (APACHE)-IV is compared to the ICU APACHE-IV score. Both scores are presented per category that predicted mortality changed, with the number of patients in each group.



Number of patients per group:

<-5%:	N=1 (0.1%)
05 – 5%:	N=1066 (76.9%)
5-10%:	N=202 (14.6%)
>10%:	N=118 (8.5%)



PART IV

CONCLUSIONS


EMERGENCY

10

GENERAL DISCUSSION

GENERAL DISCUSSION

Vital signs serve as the cornerstone of risk assessment of undifferentiated patients presenting to the Emergency Department (ED), but the interpretation of physiological variables deserves more attention. A correct interpretation of physiological variables may lead to earlier treatment (i.e., fluid resuscitation, vasopressors) and better disposition decisions, which may potentially lead to better outcomes, especially in older patients who are at higher risk for adverse outcomes.

In Chapter 1, I have described four objectives of this thesis regarding the age-adjusted interpretation of physiological variables for risk stratification in ED patients, developing a new age- and sex-adjusted risk tool for the hospital, and describing potential bias if risk tools are used for comparing the quality of care among departments. There are several key findings. First, risk tools should consider using age and sex-adjusted numerical scores instead of single cut-off values, as clinically relevant thresholds don't exist for almost all vital signs and biomarkers. Secondly, the International Early Warning Score (IEWS), a recalibrated NEWS including age and sex, substantially improves mortality prediction in the ED with better classification of patients into low and high risk. This score should replace the NEWS for better risk stratification in the ED. Thirdly, standardized mortality ratios (SMRs) are not directly comparable in patients admitted from the ED if hospitals differ in their policy to stabilize patients in the ED before ICU admission. This finding is important if the quality of care is compared among ICU departments.

10.1 Age- and sex-adjusted interpretation of physiological variables

Due to higher baseline risks, absolute mortality risk increases substantially more in older compared to younger patients with changing vital signs and biomarkers [Chapters 2 and 3]. These more marked increases in risk imply that deviating vital signs, such as a low blood pressure, have a different meaning in older patients. For example, a systolic blood pressure (SBP) of 110mmHg has a two times higher odds for mortality than an SBP above 140mmHg. Although these higher odds would still result in a low absolute mortality risk for younger patients, this blood pressure significantly impacts the outcome of elderly patients on cohort level. The same principle applies to other vital signs and most biomarkers; if deviating from their reference values, the impact is more considerable for older patients. Therefore, clinicians should be aware of these differences between young and old patients so that treatment and disposition decisions may be improved in the ED by earlier recognition of vital threads. This awareness can be created

using an age-adjusted risk tool such as the International Early Warning Score (IEWS), which we have developed and described in Chapter 8.

Points in numerical risk tools, such as early warning scores, should be based on regression coefficients, which are the logarithm of odds ratios.[175, 204] Because the odds ratios for deviating vital signs for young and old patients are comparable, the same number of points must be assigned for young and old patients for deviating vital signs, as we have done for the IEWS in Chapter 8. For deviating biomarker values, however, young and old patients do not always have similar relative risk increases. For biomarkers of renal function and homeostasis (e.g., urea, creatinine, sodium) and hemoglobin, age should be considered an effect modifier [Chapter 3]. For increasing urea and creatinine, relative mortality has more marked increases in younger than older patients. The same concept is genuine for deviating sodium and hemoglobin levels. Different reference intervals can explain the effect of age on the interpretation of these biomarkers. Both sodium and hemoglobin levels decline with increasing age in a healthy population, whereas reference intervals for urea and creatinine increase due to physiological changes.[62, 76, 77] Remarkably, most laboratories use the same reference intervals for young and old patients. Using age-adjusted reference intervals for these biomarkers may improve the interpretation of test results.[62] More importantly, the effect of age on the associations between these biomarkers and outcomes imply that these biomarkers should be used differently in risk tools for young and old patients. Different points must be assigned depending on age if these biomarkers are used for risk stratification. Because biomarkers may be expensive, not rapidly available or easily repeated, we did not consider them in the development of IEWS.

Male sex was found to be an independent predictor of mortality in different ED settings.[92-95] However, in these studies, it was not reported whether disease severity in males and females was similar at ED presentation. In Chapter 9, we have shown that despite similar characteristics and disease severity at ED arrival, male patients have higher adjusted risks for mortality and high dependency care unit admission than female patients. If healthcare providers are aware of these sex differences in clinical outcomes, they may be able to anticipate the clinical course better and provide better treatment choices for both men and women. It is therefore essential to consider these sex differences in risk stratification. For this reason, we have included sex in the newly developed IEWS.

10.2 Using thresholds for physiological variables

Many acute care guidelines and risk stratification tools use single thresholds for physiological variables, suggesting that one cut-off may discriminate between a good and bad prognosis. However, we have demonstrated in Chapters 2 and 3 that these thresholds do not exist, and mortality increases gradually with deviating vital signs or biomarkers. For mean arterial pressure (MAP), respiratory rate (RR), and temperature, we found single cut-offs at which the prognosis suddenly deteriorates. However, using these single cut-offs would ignore a further increase of risk with more extreme values, so we recommend not using single cut-offs in risk scores. Instead, risk tools should include numerical scores.

Remarkably, an SBP between 110-140mmHg is associated with increased mortality, while an SBP of 120mmHg is considered normal in a healthy population. This result shows that high blood pressure in the ED is normal, probably caused by an adrenergic response due to stress or pain. The absence of this blood pressure response could mean that the patient is critically ill. Another explanation may be that many people have hypertension and thus are used to higher SBP values. As shown in Chapter 4, the difference between the baseline SBP and initial SBP in the ED was associated with 30-day mortality. Therefore, a seemingly normal blood pressure in the ED may in fact be relative hypotension and may not be recognized as such, possibly leading to under-treatment and higher mortality, especially in older patients. Although the study described in Chapter 4 was not designed to assess whether the delta SBP was a better predictor for mortality than the initial SBP in the ED, I believe that the baseline SBP is an important variable that should be considered, if available, as the deviation from a patients' normal SBP may be relevant. In addition, we showed in Chapter 6 that tachycardia may not be helpful to detect hypotension in the ED, which emphasizes the importance to know which blood pressure value is normal for a patient.

10.3 Risk stratification tools

Risk stratification tools are commonly used for risk assessment in the ED. Early warning scores are designed for prognostication and can be used as early as in the ED and add to the clinical impression of a patient's disease severity.[182] However, most early warning scores contain methodological weaknesses in development and validation.[130] Also, they are often not sex- and age-adjusted or they use single thresholds for age in the score. In Chapter 7, we demonstrated that adding age to the National Early Warning Score (NEWS) resulted in better mortality prediction for all age categories. However, in the development of NEWS, points for vital sign disturbances

were allocated based on clinical consensus rather than on a statistical approach, which has led to discalibration.[10, 30] For this reason, we have developed the IEWS, a recalibrated sex- and age-adjusted score, which outperforms the most widely adopted NEWS.

Despite their widespread implementation, many professionals are resistant and skeptical towards formal predictions by risk scores, which is a well-known socio-psychological phenomenon.[205] Although one may argue that clinicians could assess patients' risk by experience or clinical gestalt, it is well established that even experts are biased in their perception of risk.[168, 206, 207] In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the theory of statistical prediction. Instead, they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors.[168, 208, 209] People tend to express great confidence in their biased decisions, called the illusion of validity.[210, 211] For this reason, it is crucial to use a formal risk estimate hand in hand with clinical gestalt. Furthermore, these scores were not designed to replace clinical gestalt but to delay or support it. A low-risk score may allow the clinician more time before assessing the patient, which can especially be beneficial in overcrowded EDs, whereas, for a high-risk score, clinical input may be urgently needed. The IEWS provides an objective risk that may help with clinical decision-making and add to the clinical appearance. The IEWS allows clinicians to select the patients who need an urgent clinical assessment or delay this assessment if the risk of the patient is low. Reducing alarm fatigue is especially important in EDs that become more and more overcrowded. With IEWS, fewer acute assessments are necessary, because many young patients have low risks, whereas older patients are more often at high risk for adverse outcomes. The IEWS, therefore, reduces alarm fatigue compared to NEWS by improving the recognition of patients who really need timely clinical input. This timely input can lead to earlier clinical decisions which may positively affect patients' outcome. Furthermore, it may also help to support the clinician in early initiation of comfort or palliative care if the calculated mortality risk is high in frail elderly.

Using the IEWS may lead to better treatment and disposition decisions. Consequently, this may lead to lower mortality rates and shorter hospital admission periods. Another advantage of a well-calibrated score is that it can be used for case-mix correction when comparing outcomes of patients in national and international research or quality improvement projects, which is already done with the Acute Physiology and Chronic Health Evaluation (APACHE)-IV for ICU patients. However, as we demonstrated in Chapter 9, the

calculated severity of illness and predicted mortality risk differ for a given patient if the same initial treatment is given before or after ICU admission or if the ED stay is longer, resulting in a phenomenon called lead-time bias in the literature.[37, 38] These findings imply that a correction or a margin of error should be applied whenever the quality of care among ICUs is compared using the standardized mortality ratio (SMR). Similar lead-time bias might be present in the ED if patients receive different treatments prehospital, depending on the setting. For example, if patients are more often intubated prehospital in one region because of longer distances, the physiological variables measured in the ED will be affected, and so will the predicted mortality risk. With more hospitals joining the Netherlands Emergency department Evaluation Database (NEED), quality of care can be compared among EDs in the near future. Consequently, it is essential to understand differences in the prehospital setting to correct for lead-time bias.

10.4 Limitations

The limitations of each study included in this thesis have been discussed in the relevant chapters. Some general limitations will be mentioned here. Most studies in this thesis used the NEED as the main data source. Due to the retrospective design, we cannot exclude the risk of selection and information bias. For this reason, we have validated many findings in multiple cohorts. For the development of IEWS, we tried to reduce bias by using multiple imputations for missing variables. Furthermore, although we have demonstrated that the IEWS improves mortality prediction in ED patients, the treatment paradox has likely affected our findings due to the retrospective design. This bias rests on the assumption that those at high risk receive aggressive treatment, leading to an apparent overestimation of risks because the outcome is often averted.[30] For example, fever was not associated with higher mortality, as demonstrated in Chapters 2 and 8. A better-provided quality of care, or the treatment paradox, may be partly responsible for this lower risk of patients with fever.[212] However, many other physiological explanations have been suggested that could explain the higher risk for patients with hypothermia, which should not be ignored.[213] Validation of our findings in hospital settings with almost no treatment paradox, such as rural areas in developing countries, will help to understand the treatment paradox's effect.

10.5 Future perspectives

This thesis mainly focused on investigating differences between men and women and young and old regarding the risk for adverse outcomes. More importantly, we have developed a new tool for prognostication in the ED reducing alarm fatigue and improving prognostication. The next step should

be the implementation of the IEWS in Western countries and investigate whether this will lead to better patient outcomes. This implementation may be done together with other measures that might contribute to better recognition of elderly patients at risk. For example, frailty screening instruments, such as the Clinical Frailty Score (CFS) or APOP, may help further guidance of disposition decisions in older patients.[214, 215] However, it is crucial that newly introduced risk tools or screening instruments do not lead to more administration for healthcare workers and should instead reduce workload. I believe that the biggest challenge for the future ED is implementing better risk stratification tools, such as the IEWS, without requiring physicians to calculate the score by multiple clicks or by hand. This could be achieved by implementing the scores automatically in monitors or Electronic Health Records. Prediction models can also be developed using Artificial intelligence (AI). However, so far, AI does not outperform simple logistic regression,[91] and most models are poorly reported and at high risk of bias such that the reported predictive performance is likely optimistic.[216] The advantage of a simple risk score, such as IEWS, is that clinicians understand what the predicted risk is based on. As long as acute care is delivered by humans, not by robots, the latter is essential for humans to adopt the score in clinical practice.

Although we have shown that the IEWS performs well in two Western countries, a prediction model applied in a new healthcare setting or country often produces predictions that are miscalibrated.[164, 190] This may lead to incorrect and potentially harmful clinical decisions.[189] For this reason, further validation is desired to assess the generalizability of the IEWS across multiple settings. In new healthcare settings, the IEWS might need to be updated before it can safely be applied in that new setting.[178] Instead of developing and updating the IEWS in local settings, an international dataset including data from multiple countries and healthcare systems might allow better generalisability and implementation of risk scores, such as IEWS, across different populations.[189] However, this may be a long shot due to privacy and ownership challenges. A more promising solution for better generalisability across multiple settings may be federated learning, a data-private collaborative learning method where multiple EDs could train a machine learning model on their own data at the same time, and then send their model updates to a central server to be aggregated into a consensus model.[217] In this way, data do not need to be shared among institutions. International and interdisciplinary collaborations are necessary for further improvements of prediction models used in the acute care. The NEED could contribute to this fast-growing and developing landscape.


EMERGENCY

11

SUMMARY

11.1 SUMMARY

Chapter 1 describes the background and aims of this thesis. The growing number of older patients presenting to Emergency Departments (EDs) requires better risk stratification to guide treatment and disposition decisions. Therefore, it is essential to understand the effect of age on the associations between physiological variables and outcomes.

More importantly, most risk tools are not age or sex adjusted and are not based on a statistical approach. An age and sex adjusted risk tool could improve risk stratification in the ED.

This thesis is divided into three parts and has four aims, regarding age-adjusted interpretation of physiological variables for risk stratification in ED patients, developing a new age- and sex-adjusted risk tool for the hospital, and describing potential bias if risk tools are used for comparing the quality of care among departments.

11.2 PART I

Age and sex adjusted interpretation of physiological variables in Emergency Department patients

In **Chapter 2 and 3,** we studied the associations between vital signs, biomarkers and relevant clinical outcomes in ED patients of different age categories. Data from the Netherlands Emergency department Evaluation Database (NEED) were used including patients ≥18 years.

The risk for mortality increased gradually with decreasing SBP and decreasing peripheral oxygen saturation (SpO_2). Diastolic blood pressure (DBP), mean arterial pressure (MAP) and heart rate (HR) had quasi-U-shaped associations with mortality. Mortality did not increase for temperatures anywhere in the range between 35.5-42.0°C, with a single cut-off around 35.5 °C below which mortality increased. Single cut-offs were also found for MAP <70mmHg and respiratory rate >22/min.. However, using these single cut-offs would ignore the further risk increase with more extreme values of these vital signs. For all vital signs, older patients had larger increases in absolute mortality compared to younger patients.

Mortality increased gradually for increasing C-reactive Protein (CRP), and had a U-shaped association for creatinine, sodium, leukocytes, and hemoglobin. Age significantly affected the associations between most studied blood tests and mortality.

In summary, the use of a single cut-off for each physiological variable in risk stratification tools and acute care guidelines deserves scrutiny. The association between blood tests and (adjusted) mortality depends on age. Risk tools should consider using age-adjusted numerical scores.

In **Chapter 4** we assessed how often baseline systolic blood pressure (SBP) could be retrieved from the Electronic Health Record (EHR) in older ED patients, and whether the difference between the baseline SBP and initial SBP in the ED (ΔSBP) was associated with 30-day mortality. In two Dutch EDs, baseline SBPs were found in 220 out of 300 patients (73.3%). In 72 patients with normal initial SBPs in the ED, fifteen had a negative ΔSBP with 20% mortality. A negative ΔSBP was associated with 30-day mortality (AHR 4.7; 95%-CI 1.7-12.7).

Baseline SBPs are often available in older ED patients. The ΔSBP has prognostic value and may be used as an extra variable to recognize hypotension.

Chapter 5 describes sex differences regarding disease presentations, disease severity and case-mix adjusted outcomes in the ED. The NEED was used as a data source.

Patient characteristics at ED presentation and diagnoses were comparable between sexes. In-hospital mortality was higher for males at 2.2% compared to 1.7% for females. Males had a higher adjusted risk for mortality (AOR 1.34; 95%-CI 1.24-1.46) and for high dependency care unit admission (1.46; 95%-CI 1.37-1.56). This higher risk was present for all triage categories and almost all presenting complaints.

Sex-adjusted risk scores may improve risk stratification in the ED.

In **Chapter 6** we assessed the association between SBP and heart rate (HR) in ED patients of different age categories. The NEED was used as a data source and a a cohort including patients with a suspected infection. Findings were validated in a Danish Multicenter Cohort (DMC).

No clinically relevant unadjusted or adjusted associations were found between SBP and HR in any age category or subgroup of ED patients. No increase in HR existed with a decreasing SBP during ED treatment in ED patients with a suspected infection.

In conclusion, no clinically relevant association existed between SBP and HR in ED patients of any age category, neither in subgroups nor in a different international setting. As a result, hypotension and shock cannot be recognized solely by tachycardia.

11.3 PART II

Development and external validation of an age and sex adjusted Early Warning Score

In **Chapter 7** we investigated how age affects the predictive performance of the National Early Warning Score. The DMC was used as a development cohort including 14,809 patients ≥18 years. External validation was performed in the NEED.

For NEWS alone, severe underestimation of risk was observed for persons >80 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 persons >80 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.

In summary, 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.

In **Chapter 8** we developed and externally validated the International Early Warning Score (IEWS), a recalibrated NEWS including age and sex. The NEED was used as the development cohort and the DMC as the external validation cohort. Model performance to predict in-hospital mortality was evaluated by discrimination (AUROC), calibration curves, reclassification and clinical usefulness by decision curve analysis.

The in-hospital mortality rate was 2.4% (N=2314) in the NEED and 2.5% (N=365) in the DMC. Overall, the IEWS performed significantly better than NEWS with an AUROC of 0.89 (95% confidence Intervals (CI) 0.89-0.90) versus 0.82 (CI 0.82-0.83) in the NEED and 0.87 (CI 0.85-0.88) versus 0.82 (CI 0.80-0.84) at external validation. For each age category discrimination improved with IEWS compared to NEWS. Calibration for NEWS predictions underestimated risk in older patients and overestimated risk in the youngest, while calibration improved for IEWS with a substantial improvement of reclassification of patients from low to high risk and a standardized net benefit of 5-15% in the relevant risk range for all age categories.

In conclusion, the IEWS, a recalibrated NEWS including age and sex, substantially improves mortality prediction for all ages in ED patients. The better classification into low and high risk may potentially contribute to better decision policies and the IEWS should therefore replace NEWS in the ED in Western countries.

11.4 PART III
Bias in the use of risk tools to compare quality of care

The aim of **Chapter 9** was to assess to what extent predicted mortality risk would be affected if the Acute Physiology and Chronic Health Evaluation (APACHE)-IV score was recalculated with the initial physiological variables from the ED. Secondly, to evaluate whether ED Length of Stay (LOS) was associated with a change (delta) in these APACHE-IV scores. Data from two Dutch quality registries were linked: the NEED and the National Intensive Care Evaluation (NICE) registry.

The Standardized mortality ratio (SMR) changed from 0.63 (95%CI 0.54-0.72) to 0.55 (95%CI 0.47-0.63) if ED physiological variables were considered in calculating the APACHE-IV score. ED LOS >3.9hours was associated with a slight increase in delta APACHE-IV of 1.6 (95% CI 0.4-2.8) compared to ED LOS <1.7hours.

In summary, predicted mortality risks and SMRs calculated by the APACHE IV scores are not directly comparable in patients admitted from the ED if hospitals differ in their policy to stabilize patients in the ED before ICU admission.

11.5 PART IV
General discussion

Chapter 10 constitutes the general discussion and clinical implications of the main findings, general limitations and future perspectives.


EMERGENCY

12

DUTCH SUMMARY (NEDERLANDSE SAMENVATTING)

12.1 NEDERLANDSE SAMENVATTING

Hoofdstuk 1 beschrijft de achtergrond en doelen van dit proefschrift. Vanwege het groeiende aantal oudere patiënten op de Spoedeisende Hulp (SEH) is betere risicostratificatie nodig om behandelingen en dispositie beslissingen te ondersteunen. Om die reden is het belangrijk om te begrijpen wat het effect van leeftijd is op de associatie tussen fysiologische variabelen en uitkomsten.

Belangrijk daarbij is dat bestaande risicotools niet gecorrigeerd zijn voor leeftijd of geslacht en meestal niet gebaseerd zijn op een statistische benadering. Een leeftijd en geslacht gecorrigeerde risicoscore zou risicostratificatie op de SEH kunnen verbeteren.

Dit proefschrift is verdeeld in drie delen en heeft vier doelen die betrekking hebben op leeftijd gecorrigeerde interpretatie van fysiologische variabelen voor risicostratificatie op de SEH, de ontwikkeling van een nieuw leeftijd en geslacht gecorrigeerde risicoscore, en de beschrijving van potentiële bias wanneer risicoscores gebruikt worden om kwaliteit van zorg tussen afdelingen te vergelijken.

12.2 DEEL 1
Leeftijd en geslacht gecorrigeerde interpretatie van fysiologische variabelen in Spoedeisende Hulp patiënten

In **hoofdstuk 2 en 3** onderzochten we de associatie tussen vitale parameters, biomarkers en relevante klinische uitkomsten van SEH-patiënten van verschillende leeftijdsgroepen. Data van de Netherlands Emergency department Evaluation Database (NEED) werd hiervoor gebruikt.

Het risico voor mortaliteit nam gradueel toe met afnemende systolische bloeddruk en perifere zuurstof saturatie. Diastolische bloeddruk, gemiddelde arteriële druk (MAP) en hartfrequentie hadden U-vormige associaties met mortaliteit. Het risico op overlijden nam niet toe voor lichaamstemperaturen tussen de 35.5 en 42.0 graden Celsius, terwijl er wel een plotselinge verandering was van risico onder 35.5 graden Celsius. Deze afkapwaardes, waarbij het risico plots verandert, bestonden ook voor een MAP van <70mmHg en een ademhalingsfrequentie van >22/min. Echter, het gebruik van deze afkapwaardes in risicostratificatie zou verdere toename van het risico negeren met nog extremere waardes van deze vitale parameters. Voor alle vitale parameters hadden oudere patiënten grotere toenames in het absolute risico op overlijden dan jongere patiënten.

Mortaliteit nam gradueel toe voor C-reactief proteïne (CRP), en mortaliteit had een U-vormige associatie met creatinine, natrium, leukocyten, en hemoglobine. Leeftijd had een significante invloed op de associaties tussen de meeste bloeduitslagen en mortaliteit.

Samenvattend, het gebruik van een enkele afkapwaarde voor fysiologische variabelen in risicostratificatie scores en spoedeisende geneeskunde richtlijnen verdient aandacht. De associatie tussen bloeduitslagen en (gecorrigeerde) mortaliteit is afhankelijk van leeftijd. Risicoscores zouden leeftijds gecorrigeerde numerieke scores moeten gebruiken.

In **Hoofdstuk 4** onderzochten we hoe vaak een individuele referentiewaarde voor systolische bloeddruk (SBP) achterhaald kon worden in het elektronische patiëntendossier van oudere patiënten, en of het verschil tussen deze bloeddruk en die gemeten op de SEH (ΔSBP) geassocieerd was met 30-dagen mortaliteit. Van twee Nederlandse SEH's werden referentie SBP waardes gevonden in 220 van de geïncludeerde patiënten (73.3%). Bij 72 patiënten met een normale initiële SBP op de SEH hadden 15 patiënten een negatieve ΔSBP met een mortaliteit van 20%. Een negatieve ΔSBP was geassocieerd met 30-dagen mortaliteit (Adjusted Hazard Ratio 4.7; 95% CI 1.7-12.7).

Individuele referentiewaardes voor SBP waren vaak beschikbaar bij oudere SEH-patiënten. De ΔSBP heeft prognostische waarde en kan gebruikt worden als een extra variabele in het herkennen van hypotensie.

Hoofdstuk 5 beschrijft de geslachtsverschillen in ziektepresentatie, ziekte ernst, en case-mix gecorrigeerde uitkomsten op de SEH. De NEED werd gebruikt als databron.

Patiëntkarakteristieken tijdens de SEH-presentatie en diagnoses waren vergelijkbaar tussen beide geslachten. Ziekenhuismortaliteit was hoger voor mannen met 2.2% vegeleken met 1.7% voor vrouwen. Mannen hadden een hoger gecorrigeerd risico voor mortaliteit (AOR 1.34; 95%-CI 1.24-14.6) en voor Medium Care/Intensive Care opname 1.46; 95% 1.37-1.56). Dit hogere risico was aanwezig voor alle triage categorieën en voor bijna alle ingangsklachten.

Geslacht gecorrigeerde risicoscores kunnen risicostratificatie op de SEH verbeteren.

In **Hoofdstuk 6** onderzochten we de associatie tussen SBP en hartfrequentie (HF) voor SEH-patiënten van verschillende leeftijdscategorieën. De NEED en een cohort van patiënten met een verdenking op een infectie werden gebruikt als databron. Bevindingen werden gevalideerd in een Deens multicenter cohort (DMC).

Er werd geen relevante associatie gevonden tussen bloeddruk en hartslag in verschillende leeftijdsgroepen of subgroepen van SEH-patiënten, ook niet in het DMC. Er werd geen toename in HF gevonden met een afname in bloeddruk tijdens de SEH behandeling bij patiënten met een verdenking op een infectie.

Conclusie, er bestaat geen associatie tussen SBP en HF bij SEH-patiënten van verschillende leeftijdsgroepen, ook niet in subgroepen of in een andere internationale setting. De oorzaak hiervan is mogelijk de velen confounders die een rol spelen op de SEH. Als gevolg kan een dalende bloeddruk niet herkent worden met alleen een tachycardie.

12.3 DEEL II
Ontwikkeling en externe validatie van een leeftijd en geslacht gecorrigeerde Early Warning Score

In **Hoofdstuk 7** onderzochten we het effect van leeftijd op de voorspellende waarde van de National Early Warning Score (NEWS). Het DMC werd gebruikt voor de ontwikkeling van het model, 14,809 patiënten ≥18 jaar konden geïncludeerd worden. Externe validatie van het model werd gedaan in de NEED.

Voor NEWS werd ernstige onderschatting van het risico geobserveerd bij patiënten >80 jaar. Voor de gehele populatie was de AUROC 0.82 (0.80-0.84) in het DMC en 0.75 (0.75-0.77). Wanneer NEWS werd gecombineerd met leeftijd, bestond er geen onderschatting meer van het risico in de oudere patiëntencategorie. Daarnaast nam de AUROC significant toe naar 0.86 (0.85-0.88) in het DMC en 0.82 (0.81-0.83) in de NEED.

Samenvattend, het combineren van NEWS met leeftijd verbetert de voorspellende waarde met betrekking tot ziekenhuismortaliteit, met name voor personen ouder dan 80 jaar. NEWS met leeftijd kan potentieel behandelbeslissingen en dispositie verbeteren bij aankomst van de patiënt op de SEH.

In **Hoofstuk 8** ontwikkelde we een Internationale Early Warning Score (IEWS), een gekalibreerde NEWS inclusief leeftijd en geslacht. De NEED werd gebruikt voor ontwikkeling van het model en het DMC werd gebruikt voor externe validatie. Model prestatie om ziekenhuismortaliteit te voorspellen werd geëvalueerd door discriminatie (AUROC), calibratiecurves, reclassificatie en klinische bruikbaarheid met beslissingscurves.

Ziekenhuismortaliteit was 2.4% (N=2314) in de NEED en 2.5% (N=365) in het DMC. De IEWS presteerde significant beter dan de NEWS met een AUROC van 0.89 (0.89-0.90) vergeleken met 0.82 (0.82-0.83) in de NEED, en 0.87 (0.85-0.88) vergeleken met 0.82 (0.80-0.84) in het DMC. Voor elke leeftijdscategorie verbeterde de discriminatie met IEWS vergeleken met NEWS. Calibratie van voorspellingen met NEWS onderschatte het risico in oudere patiënten en overschatte het risico van jongere patiënten. Calibratie verbeterde met IEWS en zorgde voor een substantiële verbetering van reclassificatie van patiënten in laag en hoog risico, en gestandardiseerde net benefit van 5-15% in het klinisch relevante risico bereik van alle leeftijdsgroepen.

Concluderend, de IEWS, een gekalibreerde NEWS inclusief leeftijd en geslacht, verbetert mortaliteit voorspelling substantieel voor alle leeftijden bij SEH-patiënten. Betere classificatie in laag en hoog risico kan potentieel leiden tot betere klinische beslissingen en de IEWS zou daarom de NEWS moeten vervangen op de SEH in Westerse landen.

12.4 DEEL III

Bias in het gebruik van risicoscores om kwaliteit van zorg te vergelijken

Het doel van **Hoofdstuk 9** was om te onderzoeken in welke mate de voorspelde mortaliteit beïnvloed zou worden als de Acute Physiology and Chronic Health Evaluation (APACHE)-IV score werd herberekent met de initiële fysiologische variabelen op de SEH. Daarnaast was het doel om te onderzoeken of de SEH-ligduur geassocieerd was met een verandering (delta) in deze APACHE-IV scores. Data van twee Nederlandse kwaliteitsregistraties werden gekoppeld: de NEED en de National Intensive Care Evaluation (NICE) registratie.

De gestandardizeerde mortaliteitsratio (SMR) veranderde van 0.63 (95%CI 0.54-0.72) naar 0.55 (95% 0.47-0.63) wanneer fysiologische variabelen van de SEH werden meegenomen om de APACHE-IV score te berekenen. Een SEH-ligduur van >3.9uur was geassocieerd met een lichte toename in delta APACHE-IV van 1.6 (95% CI 0.4-2.8) vergeleken met een SEH ligduur van <1.7uur.

Samenvattend, de voorspelde mortaliteitsrisico's en SMRs berekent met de APACHE-IV score zijn niet direct vergelijkbaar van patiënten opgenomen vanuit de SEH als ziekenhuizen verschillen in hun beleid om patiënten te stabiliseren op de SEH voor IC-opname.

12.5 DEEL IV
Algemene discussie

Hoofdstuk 10 bestaat uit de algemene discussie en klinische implicaties van de bevindingen, algemene limiaties en toekomstige perspectieven.


EMERGENCY

APPENDIXES

REFERENCES

1. Xu KT, Nelson BK, Berk S. The changing profile of patients who used emergency department services in the United States: 1996 to 2005. Ann Emerg Med. 2009;54(6):805-10. e7.
2. Samaras N, Chevalley T, Samaras D, Gold G. Older patients in the emergency department: a review. Ann Emerg Med. 2010;56(3):261-9.
3. Roberts DC, McKay MP, Shaffer A. Increasing rates of emergency department visits for elderly patients in the United States, 1993 to 2003. Ann Emerg Med. 2008;51(6):769-74.
4. Aminzadeh F, Dalziel WB. Older adults in the emergency department: a systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. Ann Emerg Med. 2002;39(3):238-47.
5. George G, Jell C, Todd B. Effect of population ageing on emergency department speed and efficiency: a historical perspective from a district general hospital in the UK. Emerg Med J. 2006;23(5):379-83.
6. Baum SA, Rubenstein LZ. Old people in the emergency room: Age-related differences in emergency department use and care. J Am Geriatr Soc. 1987;35(5):398-404.
7. Christakis NA. The ellipsis of prognosis in modern medical thought. Soc Sci Med. 1997;44(3):301-15.
8. Kellett J. Prognostication—the lost skill of medicine. Eur J Intern Med. 2008;19(3):155-64.
9. RoyalCollegeofPhysicians. National Early Warning Score 2. Standardizing the assessment of acute-illness severity in the NHS. National Early Warning Score (NEWS). 2012.
10. Morgan R, Lloyd-Williams F, Wright M, Morgan-Warren R. An early warning scoring system for detecting developing critical illness. 1997.
11. Ljunggren M, Castrén M, Nordberg M, Kurland L. The association between vital signs and mortality in a retrospective cohort study of an unselected emergency department population. Scand J Trauma Resusc Emerg Med. 2016;24(1):21.
12. Chester JG, Rudolph JL. Vital signs in older patients: age-related changes. J Am Med Dir Assoc. 2011;12(5):337-43.
13. Peckham H, de Gruijter NM, Raine C, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. Nature communications. 2020;11(1):1-10.
14. McGregor AJ, Beauchamp GA, Wira III CR, Perman SM, Safdar B. Sex as a biological variable in emergency medicine research and clinical practice: a brief narrative review. West J Emerg Med. 2017;18(6):1079.
15. Coventry LL, Finn J, Bremner AP. Sex differences in symptom presentation in acute myocardial infarction: a systematic review and meta-analysis. Heart Lung. 2011;40(6):477-91.
16. Reeves MJ, Bushnell CD, Howard G, et al. Sex differences in stroke: epidemiology, clinical presentation, medical care, and outcomes. The Lancet Neurology. 2008;7(10):915-26.
17. Washington DL, Bird CE. Sex differences in disease presentation in the emergency department. Ann Emerg Med. 2002;40(5):461-3.

18. Hubert H, Jaeger D, Baert V, et al. Effect of gender on out-of-hospital cardiac arrest survival: a registry-based study. Eur J Emerg Med. 2021;28(1):50-7.
19. Blom MT, Oving I, Berdowski J, van Valkengoed IG, Bardai A, Tan HL. Women have lower chances than men to be resuscitated and survive out-of-hospital cardiac arrest. Eur Heart J. 2019;40(47):3824-34.
20. Wenham C, Smith J, Morgan R. COVID-19: the gendered impacts of the outbreak. The Lancet. 2020;395(10227):846-8.
21. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. Am J Respir Crit Care Med. 2019;200(7):e45-e67.
22. Bone RC. Immunologic dissonance: a continuing evolution in our understanding of the systemic inflammatory response syndrome (SIRS) and the multiple organ dysfunction syndrome (MODS). Ann Intern Med. 1996;125(8):680-7.
23. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the Trauma Score. The Journal of trauma. 1989;29(5):623-9.
24. Shapiro NI, Wolfe RE, Moore RB, Smith E, Burdick E, Bates DW. Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule. Crit Care Med. 2003;31(3):670-5.
25. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):762-74.
26. Gilpin D, Nelson P. Revised trauma score: a triage tool in the accident and emergency department. Injury. 1991;22(1):35-7.
27. 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(1):91.
28. Spaite DW, Hu C, Bobrow BJ, et al. Mortality and prehospital blood pressure in patients with major traumatic brain injury: implications for the hypotension threshold. JAMA surgery. 2017;152(4):360-8.
29. Warmerdam M, Baris L, van Liebergen M, et al. The association between systolic blood pressure and in-hospital mortality in older emergency department patients who are hospitalised with a suspected infection. Emerg Med J. 2018:emermed-2018-207502.
30. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS—towards a national early warning score for detecting adult inpatient deterioration. Resuscitation. 2010;81(8):932-7.
31. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically III hospitalized adults. Chest. 1991;100(6):1619-36.
32. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006;34(5):1297-310.
33. Buist M, Bernard S, Nguyen TV, Moore G, Anderson J. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. Resuscitation. 2004;62(2):137-41.

34. Considine J, Jones D, Bellomo R. Emergency department rapid response systems: the case for a standardized approach to deteriorating patients. Eur J Emerg Med. 2013;20(6):375-81.
35. Flabouris A, Hart GK, George C. Outcomes of patients admitted to tertiary intensive care units after interhospital transfer: comparison with patients admitted from emergency departments. Crit Care Resusc. 2008;10(2):97-105.
36. Simpson H, Clancy M, Goldfrad C, Rowan K. Admissions to intensive care units from emergency departments: a descriptive study. Emerg Med J. 2005;22(6):423-8.
37. Dragsted L, Jörgensen J, Jensen N-H, et al. Interhospital comparisons of patient outcome from intensive care: importance of lead-time bias. Crit Care Med. 1989;17(5):418-22.
38. Tunnell R, Millar B, Smith G. The effect of lead time bias on severity of illness scoring, mortality prediction and standardised mortality ratio in intensive care—a pilot study. Anaesthesia. 1998;53(11):1045-53.
39. Rhodes A, Evans LE, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. Intensive Care Med. 2017;43(3):304-77.
40. American College of S, Committee on T. ATLS : advanced trauma life support for doctors : student course manual. Chicago, IL: American College of Surgeons; 2008.
41. Adams Jr HP, Del Zoppo G, Alberts MJ, et al. Guidelines for the early management of adults with ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups: the American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists. Stroke. 2007;38(5):1655-711.
42. Pirneskoski J, Lääperi M, Kuisma M, Olkkola KT, Nurmi J. Ability of prehospital NEWS to predict 1-day and 7-day mortality is reduced in the older adult patients. Emerg Med J. 2021.
43. Cretikos MA, Bellomo R, Hillman K, Chen J, Finfer S, Flabouris A. Respiratory rate: the neglected vital sign. Med J Aust. 2008;188(11):657-9.
44. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007;165(6):710-8.
45. van der Veen D, Heringhaus C, de Groot B. Appropriateness, Reasons and Independent Predictors of Consultations in the Emergency Department (ED) of a Dutch Tertiary Care Center: A Prospective Cohort Study. PLoS One. 2016;11(2):e0149079.
46. Woods RA, Lee R, Ospina MB, et al. Consultation outcomes in the emergency department: exploring rates and complexity. Canadian Journal of Emergency Medicine. 2008;10(1):25-31.
47. Rubulotta F, Marshall JC, Ramsay G, Nelson D, Levy M, Williams M. Predisposition, insult/infection, response, and organ dysfunction: A new model for staging severe sepsis. Crit Care Med. 2009;37(4):1329-35.
48. Lim WS, Van der Eerden M, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58(5):377-82.

49. Klang E, Soffer S, Shimon Shahar M, et al. Association of normal systolic blood pressure in the emergency department with higher in-hospital mortality among hypertensive patients. The Journal of Clinical Hypertension. 2019;21(12):1841-8.
50. Kristensen AKB, Holler JG, Mikkelsen S, Hallas J, Lassen A. Systolic blood pressure and short-term mortality in the emergency department and prehospital setting: a hospital-based cohort study. Critical Care. 2015;19(1):158.
51. Alam N, Hobbelink EL, van Tienhoven A-J, van de Ven PM, Jansma EP, Nanayakkara PW. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. Resuscitation. 2014;85(5):587-94.
52. Morgan R, Wright M. In defence of early warning scores. Br J Anaesth. 2007;99(5):747-8.
53. Oyetunji TA, Chang DC, Crompton JG, et al. Redefining hypotension in the elderly: normotension is not reassuring. Arch Surg. 2011;146(7):865-9.
54. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818-29.
55. de Groot B, de Deckere ER, Flameling R, Sandel MH, Vis A. Performance of illness severity scores to guide disposition of emergency department patients with severe sepsis or septic shock. Eur J Emerg Med. 2012;19(5):316-22.
56. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. JAMA. 1993;270(20):2478-86.
57. Howell MD, Talmor D, Schuetz P, Hunziker S, Jones AE, Shapiro NI. Proof of principle: the predisposition, infection, response, organ failure sepsis staging system. Crit Care Med. 2011;39(2):322-7.
58. Hall M, Bebb OJ, Dondo TB, et al. Guideline-indicated treatments and diagnostics, GRACE risk score, and survival for non-ST elevation myocardial infarction. Eur Heart J. 2018;39(42):3798-806.
59. Leppäniemi A, Tolonen M, Tarasconi A, et al. 2019 WSES guidelines for the management of severe acute pancreatitis. World J Emerg Surg. 2019;14(1):27.
60. van Bemmel T, Woittiez K, Blauw GJ, et al. Prospective study of the effect of blood pressure on renal function in old age: the Leiden 85-Plus Study. 2006;17(9):2561-6.
61. Epstein M. Aging and the kidney. J Am Soc Nephrol. 1996;7(8):1106.
62. den Elzen WP, Brouwer N, Thelen MH, et al. NUMBER: standardized reference intervals in the Netherlands using a 'big data'approach. 2018;57(1):42-56.
63. Raven W, van den Hoven EM, Gaakeer MI, et al. The association between presenting complaints and clinical outcomes in emergency department patients of different age categories. European Journal of Emergency Medicine: Official Journal of the European Society for Emergency Medicine. 2021.
64. Wood SN. Generalized additive models: an introduction with R: CRC press; 2017.
65. Mohammed MA, Rudge G, Wood G, et al. Which is more useful in predicting hospital mortality-dichotomised blood test results or actual test values? A retrospective study in two hospitals. PLoS One. 2012;7(10):e46860.
66. Pedersen PB, Henriksen DP, Brabrand M, Lassen AT. Level of vital and laboratory values on arrival, and increased risk of 7-day mortality among adult patients in the emergency department: a population-based cohort study. BMJ Open. 2020;10(11):e038516.

67. Asadollahi K, Hastings IM, Gill GV, Beeching NJ. Prediction of hospital mortality from admission laboratory data and patient age: a simple model. Emerg Med Australas. 2011;23(3):354-63.
68. Froom P, Shimoni Z. Prediction of hospital mortality rates by admission laboratory tests. Clin Chem. 2006;52(2):325-8.
69. Tabak YP, Sun X, Nunez CM, Johannes RS. Using electronic health record data to develop inpatient mortality predictive model: Acute Laboratory Risk of Mortality Score (ALaRMS). J Am Med Inform Assoc. 2014;21(3):455-63.
70. Kristensen M, Iversen AKS, Gerds TA, et al. Routine blood tests are associated with short term mortality and can improve emergency department triage: a cohort study of> 12,000 patients. Scand J Trauma Resusc Emerg Med. 2017;25(1):115.
71. Prytherch D, Sirl J, Schmidt P, Featherstone PI, Weaver P, Smith GB. The use of routine laboratory data to predict in-hospital death in medical admissions. Resuscitation. 2005;66(2):203-7.
72. Redfern OC, Pimentel MA, 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.
73. Jarvis SW, Kovacs C, Badriyah T, et al. Development and validation of a decision tree early warning score based on routine laboratory test results for the discrimination of hospital mortality in emergency medical admissions. Resuscitation. 2013;84(11):1494-9.
74. Brabrand M, Knudsen T, Hallas J. Identifying admitted patients at risk of dying: a prospective observational validation of four biochemical scoring systems. BMJ open. 2013;3(6).
75. Silke B, Kellett J, Rooney T, Bennett K, O'riordan D. An improved medical admissions risk system using multivariable fractional polynomial logistic regression modelling. QJM: An International Journal of Medicine. 2010;103(1):23-32.
76. Lindner G, Pfortmüller CA, Leichtle AB, Fiedler GM, Exadaktylos AK. Age-related variety in electrolyte levels and prevalence of dysnatremias and dyskalemias in patients presenting to the emergency department. Gerontology. 2014;60(5):420-3.
77. Nilsson-Ehle H, Jagenburg R, Landahl S, Svanborg A. Blood haemoglobin declines in the elderly: implications for reference intervals fromage 70 to 88. Eur J Haematol. 2000;65(5):297-305.
78. Roenhoej R, Hasselbalch RB, Schultz M, et al. Abnormal routine blood tests as predictors of mortality in acutely admitted patients. Clin Biochem. 2020;77:14-9.
79. Lloyd-Sherlock P, Beard J, Minicuci N, Ebrahim S, Chatterji S. Hypertension among older adults in low-and middle-income countries: prevalence, awareness and control. International journal of epidemiology. 2014;43(1):116-28.
80. Kaplan MS, Huguet N, Feeny DH, McFarland BH. Self-reported hypertension prevalence and income among older adults in Canada and the United States. Social science & medicine. 2010;70(6):844-9.
81. Kalra D. Medicine in Europe: Electronic health records: the European scene. BMJ. 1994;309(6965):1358-61.

82. Candel BG, Duijzer R, Gaakeer MI, de Jonge E, de Groot B. The association between vital signs and relevant clinical outcomes in emergency department patients of different age categories. Emerg Med J. 2022;39(12):903-11.
83. Futier E, Lefrant J-Y, Guinot P-G, et al. Effect of individualized vs standard blood pressure management strategies on postoperative organ dysfunction among high-risk patients undergoing major surgery: a randomized clinical trial. JAMA. 2017;318(14):1346-57.
84. Asfar P, Meziani F, Hamel J-F, et al. High versus low blood-pressure target in patients with septic shock. N Engl J Med. 2014;370:1583-93.
85. Bruijns SR, Guly HR, Bouamra O, Lecky F, Wallis LA. The value of the difference between ED and prehospital vital signs in predicting outcome in trauma. Emerg Med J. 2014;31(7):579-82.
86. Kamata K, Abe T, Aoki M, Deshpande G, Saitoh D, Tokuda Y. Dynamic vital signs may predict in-hospital mortality in elderly trauma patients. Medicine (Baltimore). 2020;99(25):e20741.
87. Arbabi S, Jurkovich GJ, Wahl WL, et al. A comparison of prehospital and hospital data in trauma patients. Journal of Trauma and Acute Care Surgery. 2004;56(5):1029-32.
88. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. Resuscitation. 2016;102:1-5.
89. Brekke IJ, Puntervoll LH, Pedersen PB, Kellett J, Brabrand M. The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. PLoS One. 2019;14(1).
90. Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. J Am Med Inform Assoc. 2010;17(3):288-94.
91. 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. 2021:104496.
92. Teubner DJ, Considine J, Hakendorf P, Kim S, Bersten AD. Model to predict inpatient mortality from information gathered at presentation to an emergency department: The Triage Information Mortality Model (TIMM). Emerg Med Australas. 2015;27(4):300-6.
93. Xie F, Liu N, Wu SX, et al. Novel model for predicting inpatient mortality after emergency admission to hospital in Singapore: retrospective observational study. BMJ open. 2019;9(9):e031382.
94. Beglinger B, Rohacek M, Ackermann S, et al. Physician's first clinical impression of emergency department patients with nonspecific complaints is associated with morbidity and mortality. Medicine. 2015;94(7).
95. Ha DT, Dang TQ, Tran NV, Pham TN, Nguyen ND, Nguyen TV. Development and validation of a prognostic model for predicting 30-day mortality risk in medical patients in emergency department (ED). Sci Rep. 2017;7(1):1-10.
96. Safdar B, Ona Ayala KE, Ali SS, et al. Inclusion of sex and gender in Emergency Medicine Research—A 2018 update. Acad Emerg Med. 2019;26(3):293-302.
97. van der Veen D, Remeijer C, Fogteloo AJ, Heringhaus C, de Groot B. Independent determinants of prolonged emergency department length of stay in a tertiary care centre: a prospective cohort study. Scand J Trauma Resusc Emerg Med. 2018;26(1):81.
98. OECDHealthStatistics. Eurostat Database. 2020.
99. Barba R, Marco J, Canora J, et al. Prolonged length of stay in hospitalized internal medicine patients. Eur J Intern Med. 2015;26(10):772-5.

100. Angele MK, Pratschke S, Hubbard WJ, Chaudry IH. Gender differences in sepsis: cardiovascular and immunological aspects. Virulence. 2014;5(1):12-9.
101. Labiche LA, Chan W, Saldin KR, Morgenstern LB. Sex and acute stroke presentation. Ann Emerg Med. 2002;40(5):453-60.
102. El-Menyar A, Zubaid M, Sulaiman K, et al. Atypical presentation of acute coronary syndrome: a significant independent predictor of in-hospital mortality. J Cardiol. 2011;57(2):165-71.
103. Sandberg K, Ji H. Sex differences in primary hypertension. Biol Sex Differ. 2012;3(1):1-21.
104. Backus BE, Six AJ, Kelder JC, et al. Chest pain in the emergency room: a multicenter validation of the HEART Score. Crit Pathw Cardiol. 2010;9(3):164-9.
105. Karounos M, Chang AM, Robey JL, et al. TIMI risk score: does it work equally well in both males and females? Emerg Med J. 2007;24(7):471-4.
106. Bank IE, de Hoog VC, de Kleijn DP, et al. Sex-based differences in the performance of the HEART score in patients presenting to the emergency department with acute chest pain. Journal of the American Heart Association. 2017;6(6):e005373.
107. Hinson JS, Martinez DA, Cabral S, et al. Triage performance in emergency medicine: a systematic review. Ann Emerg Med. 2019;74(1):140-52.
108. Hall JE, Hall ME. Guyton and Hall textbook of medical physiology e-Book, Ch24, 293-301: Elsevier Health Sciences; 2020.
109. Kasper D, Fauci A, Hauser S, Longo D, Jameson J, Loscalzo J. Harrison's principles of internal medicine, 19e, Chapter 232: Basic Biology of the Cardiovascular System.: Mcgraw-hill; 2015.
110. Subcommittee A, Group IAW. Advanced trauma life support (ATLS®): the ninth edition. The journal of trauma and acute care surgery. 2013;74(5):1363-6.
111. Mutschler M, Nienaber U, Brockamp T, et al. A critical reappraisal of the ATLS classification of hypovolaemic shock: does it really reflect clinical reality? Resuscitation. 2013;84(3):309-13.
112. Brasel KJ, Guse C, Gentilello LM, Nirula R. Heart rate: is it truly a vital sign? Journal of Trauma and Acute Care Surgery. 2007;62(4):812-7.
113. Guly H, Bouamra O, Little R, et al. Testing the validity of the ATLS classification of hypovolaemic shock. Resuscitation. 2010;81(9):1142-7.
114. Victorino GP, Battistella FD, Wisner DH. Does tachycardia correlate with hypotension after trauma? J Am Coll Surg. 2003;196(5):679-84.
115. Ahmad S, Tejuja A, Newman KD, Zarychanski R, Seely AJ. Clinical review: a review and analysis of heart rate variability and the diagnosis and prognosis of infection. Critical Care. 2009;13(6):1-7.
116. Griffin MP, O'Shea TM, Bissonette EA, Harrell FE, Lake DE, Moorman JR. Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness. Pediatr Res. 2003;53(6):920-6.
117. Valentini M, Parati G. Variables influencing heart rate. Prog Cardiovasc Dis. 2009;52(1):11-9.
118. Brabrand M, Hallas P, Hansen SN, Jensen KM, Madsen JLB, Posth S. 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;45:32-6.

119. Nissen SK, Candel BG, Nickel CH, et al. The Impact of Age on Predictive Performance of National Early Warning Score at Arrival to Emergency Departments: Development and External Validation. Ann Emerg Med. 2022;79(4):354-63.
120. de Groot B, Ansems A, Gerling DH, et al. The association between time to antibiotics and relevant clinical outcomes in emergency department patients with various stages of sepsis: a prospective multi-center study. Critical care. 2015;19(1):1-12.
121. Goldman L, Schafer AI. Goldman-Cecil Medicine, 26th. Chapter 8: Cardiovascular disease. : Elsevier; 2020.
122. Secher NH, Bie P. Bradycardia during reversible haemorrhagic shock-a forgotten observation? Clin Physiol. 1985;5(4):315-23.
123. Oberg B, Thoren P. Increased activity in vagal cardiac afferents correlated to the appearance of reflex bradycardia during severe hemorrhage in cats. Acta Physiol Scand. 1970;80(4):22A-3A.
124. Castenfors J, Sjöstrand T. Circulatory Control via Vagal Afferents: I. Adjustment of heart rate to variations of blood volume in the rat. Acta Physiol Scand. 1972;84(3):347-54.
125. Secher N, Werner C, Warberg J, Bie P. Bradycardia during severe but reversible hypovolemic shock in man. Circ Shock. 1984;14(4):267-74.
126. Sander-Jensen K, Secher N, Bie P, Warberg J, Schwartz T. Vagal slowing of the heart during haemorrhage: observations from 20 consecutive hypotensive patients. Br Med J (Clin Res Ed). 1986;292(6517):364-6.
127. Barriot P, Riou B. Hemorrhagic shock with paradoxical bradycardia. Intensive Care Med. 1987;13(3):203-7.
128. Morelli A, D'Egidio A, Passariello M. Tachycardia in septic shock: pathophysiological implications and pharmacological treatment. Annual Update in Intensive Care and Emergency Medicine 2015: Springer; 2015. p. 115-28.
129. Parrillo JE. Pathogenetic mechanisms of septic shock. N Engl J Med. 1993;328(20):1471-7.
130. 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.
131. Physicians RCo. National Early Warning Score (NEWS) 2: standardising the assessment of acute-illness severity in the NHS. Updated report of a working party 2017. 2021.
132. RCoPo L, editor National early warning score (NEWS): standardising the assessment of acute-illness severity in the NHS—report of a working party. Royal College of Physicians; 2012.
133. 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(2):109-15.
134. 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(5):453.
135. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM. 2001;94(10):521-6.
136. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Journal of British Surgery. 2015;102(3):148-58.

137. Brabrand M, Knudsen T, Hallas J, Graham CA, Kellett J. The Paris score can reliably predict 7-day all-cause mortality for both acute medical and surgical patients: an international validation study. QJM: An International Journal of Medicine. 2018;111(10):721-5.
138. Nielsen MK, Henriksen DP, Østergaard KB, Dahlin J, Brabrand M. Association between presenting complaints of acutely admitted medical patients and mortality: a cohort study. Eur J Intern Med. 2018;54:e29-e32.
139. Öhman M, Atkins T, Cooksley T, Brabrand M. Validation of the MARS: a combined physiological and laboratory risk prediction tool for 5-to 7-day in-hospital mortality. QJM: An International Journal of Medicine. 2018;111(6):385-8.
140. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol. 2014;29(8):541-9.
141. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol. 2015;7:449.
142. Harrell FE. Regression Modeling Strategies. vol 45. Springer International Publishing, Cham2015.
143. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. Journal of statistical software. 2011;45:1-67.
144. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581-92.
145. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016;74:167-76.
146. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352.
147. Ensor J, Martin E, Riley R. Pmsampsize: calculates the minimum sample size required for developing a multivariable prediction model. 2019.
148. Heymans M, van Buuren S, Knol D, van Mechelen W, de Vet H. CW (2007). Variable selection under multiple imputation using the bootstrap in a prognostic study. BMC Med Res Methodol.7:33.
149. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338.
150. Collins GS, Moons KG. Comparing risk prediction models. British Medical Journal Publishing Group; 2012.
151. 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;47:222-6.
152. de Groot B, Lameijer J, de Deckere ER, Vis A. 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(4):292-300.
153. 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. Prehosp Emerg Care. 2017;21(6):734-43.

154. 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):1-9.
155. 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.
156. 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(3):291-300.
157. 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. 2022;13(2):453-61.
158. Dent E, Martin FC, Bergman H, Woo J, Romero-Ortuno R, Walston JD. Management of frailty: opportunities, challenges, and future directions. The Lancet. 2019;394(10206):1376-86.
159. Groenwold RH, Moons KG, 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.
160. 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(1):57-69.
161. Brabrand M, Kellett J, Opio M, Cooksley T, Nickel C. Should impaired mobility on presentation be a vital sign? Acta Anaesthesiol Scand. 2018;62(7):945-52.
162. Hillman K. Predicting outcome in the seriously ill-A new approach. 2018. p. 878-9.
163. Pencina MJ, D'Agostino Sr RB, D'Agostino Jr 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(2):157-72.
164. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):1-7.
165. 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.
166. 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(11):e033301.
167. O'Neill S, 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(1):1-19.
168. Kahneman D, Tversky A. On the psychology of prediction. Psychol Rev. 1973;80(4):237.
169. Smith MB, Chiovaro JC, O'Neil M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. Annals of the American Thoracic Society. 2014;11(9):1454-65.
170. McGinley A, Pearse RM. A national early warning score for acutely ill patients. British Medical Journal Publishing Group; 2012.

171. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. Resuscitation. 2013;84(4):465-70.
172. Candel BG, Dap S, Raven W, et al. Sex differences in clinical presentation and risk stratification in the Emergency Department: An observational multicenter cohort study. Eur J Intern Med. 2022;95:74-9.
173. Raisi-Estabragh Z, Kobo O, Elbadawi A, et al. Differential Patterns and Outcomes of 20.6 Million Cardiovascular Emergency Department Encounters for Men and Women in the United States. Journal of the American Heart Association. 2022:e026432.
174. Harrell FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer; 2015.
175. Moons KG, Harrell FE, Steyerberg EW. Should scoring rules be based on odds ratios or regression coefficients? J Clin Epidemiol. 2002;55(10):1054-5.
176. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology (Cambridge, Mass). 2010;21(1):128.
177. Panday RN, Minderhoud T, Alam N, Nanayakkara P. Prognostic value of early warning scores in the emergency department (ED) and acute medical unit (AMU): A narrative review. Eur J Intern Med. 2017;45:20-31.
178. Steyerberg EW. Clinical prediction models: Springer; 2019.
179. Vergouwe Y, Royston P, Moons KG, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. J Clin Epidemiol. 2010;63(2):205-14.
180. Tarassenko L, Clifton DA, Pinsky MR, Hravnak MT, Woods JR, Watkinson PJ. Centile-based early warning scores derived from statistical distributions of vital signs. Resuscitation. 2011;82(8):1013-8.
181. Watkinson PJ, Pimentel MAF, Clifton DA, Tarassenko L. Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data. Resuscitation. 2018;129:55-60.
182. Bingisser R, Baerlocher SM, Kuster T, Nieves Ortega R, Nickel CH. Physicians' disease severity ratings are non-inferior to the emergency severity index. Journal of Clinical Medicine. 2020;9(3):762.
183. Brabrand M, Hallas J, Knudsen T. Nurses and physicians in a medical admission unit can accurately predict mortality of acutely admitted patients: a prospective cohort study. PLoS One. 2014;9(7):e101739.
184. 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(1):e024636.
185. Elliott A, Taub N, Banerjee J, et al. Does the Clinical Frailty Scale at triage predict outcomes from emergency care for older people? Ann Emerg Med. 2021;77(6):620-7.
186. Gilbert T, Neuburger J, Kraindler J, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. The Lancet. 2018;391(10132):1775-82.
187. Elliott A, Phelps K, Regen E, Conroy SP. Identifying frailty in the Emergency Department—Feasibility study. Age Ageing. 2017;46(5):840-5.

188. Candel BG, Raven W, Lameijer H, et al. The effect of treatment and clinical course during Emergency Department stay on severity scoring and predicted mortality risk in Intensive Care patients. Critical Care. 2022;26(1):1-9.
189. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353.
190. Steyerberg EW, Nieboer D, Debray TP, van Houwelingen HC. Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: an overview and illustration. Stat Med. 2019;38(22):4290-309.
191. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol. 2019;110:63-73.
192. Churpek MM, Yuen TC, Edelson DP. Predicting clinical deterioration in the hospital: the impact of outcome selection. Resuscitation. 2013;84(5):564-8.
193. Pimentel MA, Redfern OC, Gerry S, et al. A comparison of the ability of the National Early Warning Score and the National Early Warning Score 2 to identify patients at risk of in-hospital mortality: a multi-centre database study. Resuscitation. 2019;134:147-56.
194. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011;20(1):40-9.
195. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol. 2003;56(5):441-7.
196. Bouch DC, Thompson JP. Severity scoring systems in the critically ill. Continuing education in anaesthesia, critical care & pain. 2008;8(5):181-5.
197. van de Klundert N, Holman R, Dongelmans DA, de Keizer NF. Data resource profile: the Dutch National Intensive Care Evaluation (NICE) registry of admissions to adult intensive care units. Int J Epidemiol. 2015;44(6):1850-h.
198. Arts D, de Keizer N, Scheffer G-J, de Jonge E. Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry. Intensive Care Med. 2002;28(5):656-9.
199. Koetsier A, Peek N, De Keizer N. Identifying types and causes of errors in mortality data in a clinical registry using multiple information systems. Quality of Life through Quality of Information: IOS Press; 2012. p. 771-5.
200. Zeserson E, Goodgame B, Hess JD, et al. Correlation of venous blood gas and pulse oximetry with arterial blood gas in the undifferentiated critically ill patient. J Intensive Care Med. 2018;33(3):176-81.
201. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med. 2001;345(19):1368-77.
202. Pappachan JV, Millar B, Bennett ED, Smith GB. Comparison of outcome from intensive care admission after adjustment for case mix by the APACHE III prognostic system. Chest. 1999;115(3):802-10.
203. Gunnerson KJ, Bassin BS, Havey RA, et al. Association of an emergency department–based intensive care unit with survival and inpatient intensive care unit admissions. JAMA network open. 2019;2(7):e197584-e.
204. Harrell F. Regression coefficients and scoring rules. J Clin Epidemiol. 1996;49(7):819.

205. Grove WM, Meehl PE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. Psychol Public Policy Law. 1996;2(2):293.
206. van der Vegt AE, Holman M, ter Maaten JC. The value of the clinical impression in recognizing and treating sepsis patients in the emergency department. Eur J Emerg Med. 2012;19(6):373-8.
207. Kahneman D, Klein G. Conditions for intuitive expertise: a failure to disagree. Am Psychol. 2009;64(6):515.
208. Kahneman D, Tversky A. Subjective probability: A judgment of representativeness. Cogn Psychol. 1972;3(3):430-54.
209. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. Cogn Psychol. 1973;5(2):207-32.
210. Einhorn HJ, Hogarth RM. Confidence in judgment: Persistence of the illusion of validity. Psychol Rev. 1978;85(5):395.
211. Bushyhead JB, Christensen-Szalanski JJ. Feedback and the illusion of validity in a medical clinic. Med Decis Making. 1981;1(2):115-22.
212. Inghammar M, Sunden-Cullberg J. Prognostic significance of body temperature in the emergency department vs the ICU in Patients with severe sepsis or septic shock: A nationwide cohort study. PLoS One. 2020;15(12):e0243990.
213. Shimazui T, Nakada T-a, Walley KR, et al. Significance of body temperature in elderly patients with sepsis. Critical Care. 2020;24(1):1-9.
214. de Gelder J, Lucke JA, Blomaard L, et al. Optimization of the APOP screener to predict functional decline or mortality in older emergency department patients: cross-validation in four prospective cohorts. Exp Gerontol. 2018;110:253-9.
215. Blomaard LC, Lucke JA, Gelder Jd, et al. The APOP screener and clinical outcomes in older hospitalized internal medicine patients. Neth J Med. 2020;78(1):25-33.
216. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020;369.
217. Sheller MJ, Edwards B, Reina GA, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep. 2020;10(1):1-12.

LIST OF PUBLICATIONS

This thesis

Candel, B. G., Dap, S., Raven, W., Lameijer, H., Gaakeer, M. I., de Jonge, E., & de Groot, B. (2022). Sex differences in clinical presentation and risk stratification in the Emergency Department: An observational multicenter cohort study. *European Journal of Internal Medicine*, *95*, 74-79.

Candel, B. G., Duijzer, R., Gaakeer, M. I., Ter Avest, E., Sir, Ö., Lameijer, H., ... & de Groot, B. (2022). The association between vital signs and clinical outcomes in emergency department patients of different age categories. *Emergency Medicine Journal*, 39(12), 903-911.

Candel, B. G., Khoudja, J., Gaakeer, M. I., Ter Avest, E., Lameijer, H., Hessels, R. A., ... & de Groot, B. (2022). Age-adjusted interpretation of biomarkers of renal function and homeostasis, inflammation, and circulation in Emergency Department patients. *Scientific reports*, *12*(1), 1-9.

Nissen, S. K., **Candel, B. G**., Nickel, C. H., de Jonge, E., Ryg, J., Bogh, S. 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.

Candel, B. G., van Ingen, I. B., van Doormalen, I. P., Raven, W., Mignot-Evers, L. A., de Jonge, E., & de Groot, B. (2022). The difference between the patients' initial and previously measured systolic blood pressure as predictor of mortality in older emergency department patients. *European Geriatric Medicine*, *13*(2), 359-365.

Candel, B. G., Raven, W., Lameijer, H., Thijssen, W. A., Temorshuizen, F., Boerma, C., ... & de Groot, B. (2022). The effect of treatment and clinical course during Emergency Department stay on severity scoring and predicted mortality risk in Intensive Care patients. *Critical Care*, *26*(1), 1-9.

Candel, B. G., Nissen, S. K., Nickel, C. H., Raven, W., Thijssen, W., Gaakeer, M. I., ... & de Groot, B. (2023). Development and External Validation of the International Early Warning Score for Improved Age-and Sex-Adjusted In-Hospital Mortality Prediction in the Emergency Department. *Critical Care Medicine*, *51*(7), 881-891.

Candel, B. G., Raven, W., Nissen, S. K., Morsink, M. E., Gaakeer, M. I., Brabrand, M., ... & de Groot, B. (2023). The association between systolic blood pressure and heart rate in Emergency Department Patients: A multicenter cohort study. *The Journal of Emergency Medicine*.

Other publications

Candel, B. G., Admiraal-van de Pas, Y., & Smit-van de Wiel, F. (2020). Suspicion of abdominal injuries in high-energy trauma patients: which clinical factors influence decision making for diagnostic imaging?. *Acta Chirurgica Belgica*, *120*(4), 223-230.

Candel, B. G., Vaes, S. M., van Bree, E. M., Groenen, S. M., Derkx, F., Mignot, L. A., & de Groot, B. (2021). Telemedicine in the emergency department to decrease personal protective equipment use: a before-and-after study. *Emergency Medicine Journal*, *38*(3), 224-228.

Candel, B. G., Karsonopoero, C. A., Arpots, R., Mans, S., Morsink, M. E., & Duijst, W. L. (2022). Nauwelijks bezwaar van SEH-patiënten tegen terugkoppeling medische gegevens. *Nederlands Tijdschrift voor Geneeskunde*, 166-D6384.

Candel, B. G., & Nissen, S. K. (2023). Use TRIPOD when validating clinical prediction models. *Emergency Medicine Journal, 40*(1), 81-81.

Candel, B. G., de Groot, B., Nissen, S. K., Thijssen, W. A., Lameijer, H., & Kellett, J. (2022). The prediction of 24-h mortality by the respiratory rate and oxygenation index compared with National Early Warning Score in emergency department patients: an observational study. *European Journal of Emergency Medicine*, 10-1097.

Kellett, J., Holland, M., & **Candel, B. G.** (2023). Using Vital Signs to Place Acutely Ill Patients Quickly and Easily into Clinically Helpful Pathophysiologic Categories: Derivation and Validation of Eight Pathophysiologic Categories in Two Distinct Patient Populations of Acutely Ill Patients. *The Journal of Emergency Medicine*, *64*(2), 136-144.

CURRICULUM VITAE

Bart Gerard Jan Candel was born on the 29th of November 1990 in Oosterhout, the Netherlands. After graduating from secondary school at the Frencken College in Oosterhout he studied medicine at the University of Maastricht. He took a gap year between his bachelor's and master's and was living for half a year in Cape Town following an internship in trauma surgery at the Tygerberg Hospital in 2011. Bart was an active member of the athletics association UROS, triathlon association Ferro Mosae, and student association KoKo in Maastricht.

After his studies, he started to gain interest in acute care and worked as a resident in Emergency Medicine and Intensive Care Medicine at the Máxima Medical Center in Veldhoven, Catharina Hospital in Eindhoven, dr. Horacio Oduber Hospital in Aruba, and the Elisabeth-Tweesteden hospital in Tilburg. Formal training in Emergency Medicine started in 2019 under the supervision of Mieke van Campen-Hoekstra, followed by Fenneke van de Wiel. During his residency, Bart started doing research in the field of Emergency Medicine. Around 2020 he started his Ph.D. under the supervision of Dr. Bas de Groot and Prof. Dr. Evert de Jonge at the Leiden University Medical Center, studying age and sex adjusted risk stratification in the acute care, which has resulted in this thesis. Since 2022 he is an Emergency Physician and he has worked at the Emergency Department of the St. John of God Midland Hospital, and Fiona Stanley Hospital in Perth. Bart will continue doing research in the field of Emergency Medicine.

DANKWOORD

De totstandkoming van dit proefschrift is het resultaat van vele uren werk naast de opleiding tot Spoedeisende Hulp-arts. Dit proefschrift was echter nooit tot stand gekomen zonder de hulp, steun en vertrouwen van velen. Graag zou ik een aantal van hen alhier hartelijk willen bedanken.

Allereerst wil ik mijn dank uitspreken aan mijn promotor Prof. dr. E. de Jonge en copromotor dr. B. de Groot. Beste Evert, ondanks de hectische Covid periode waarbij jouw aandacht, als hoofd van de afdeling, op de Intensive Care gericht was, heb ik jouw begeleiding als zeer secuur, kritisch en vlot ervaren. Er was voldoende ruimte voor discussie via Zoom of Teams waardoor het onderzoek naar een hoger niveau getild kon worden. Heel erg bedankt voor de afgelopen tijd. Beste Bas, jouw enthousiasme heeft mij vanaf dag 1 ook enthousiast gemaakt voor de wetenschap en gemotiveerd om de nodige uurtjes in de verschillende studies te steken. Voor mij als jonge arts ben je een voorbeeldfunctie en inspiratiebron binnen de Spoedeisende Geneeskunde en ik hoop dat we nog een lange tijd samen mogen werken in de toekomst.

Daarnaast wil ik graag mijn broers en tevens paranimfen, Jasper en Jeroen Candel, bedanken. Jeroen, niet alleen jouw carriere als associate professor binnen landbouw en voedselbeleid aan de Universiteit van Wageningen is bewonderenswaardig, ook jouw gedrevenheid voor hobby's zoals vogels spotten en wandelen is fantastisch en herken ik bij mezelf en andere Candels. Ik hoop dat we nog een lange tijd samen activiteiten kunnen ondernemen zoals ski of wandeltripjes. Jasper, als eeneiige tweelingbroer hebben wij vaak niet veel woorden nodig om elkaar te begrijpen. Het leven is een stuk makkelijker als je een tweelingboer hebt gezien je altijd iemand hebt om op terug te vallen, een soort externe harde schijf. Als meest idealistische persoon van ons drie heb je vaak rake adviezen die ik goed heb kunnen gebruiken tijdens mijn promotietraject. Ik ben blij dat jullie mijn broers zijn.

Pap, mam, fantastisch hoe jullie mij vrij lieten als kind en adolescent zodat ik me kon ontplooien op school, met muziek en sport. Juist doordat ik het leven zelf mocht en kon ontdekken, leerde ik om fouten te maken en hiervan te leren (of niet) en mezelf te verbeteren. Bedankt voor jullie vertrouwen, trots en liefde door al die jaren heen. Pap, hoewel ik jouw perfectionisme niet van je geërfd heb, herken ik veel van mijn karaktereigenschappen in jou. Bedankt voor je steun, je bent een geweldige vader. Mam, als pelgrim weet jij als geen ander wat doorzettingsvermogen is. Geweldig om te zien hoe jij in het leven staat. Je bent een geweldige moeder.

Opa, als 93-jarige man zit je vol met wijsheid, kennis en ervaring waar weinig mensen aan kunnen tippen. Ik ben blij met de bijzondere band die we hebben en ik wil je bedanken voor je steun en actieve betrokkenheid bij ons leven.

Vrienden waren van essentieel belang voor het schrijven van dit proefschrift. Door de jaren heen heb ik een fantastische mooie groep vrienden opgegbouwd waar ik erg dankbaar voor ben. Met onder andere de vrienden van de studententijd: Jaarclub Quincy, dispuut Festglede, vrienden van triathlonvereniging Ferro Mosae en atletiekvereniging Uros, als ook mijn vrienden uit Eindhoven. We kunnen goed met elkaar lachen en ik vind het geweldig dat ik met velen erg goed contact heb. Bedankt voor de vele uurtjes samen sporten, terrasjes en uurtjes in de kroeg.

Dan als laatste, lieve Oana, jij bent de belangrijkste persoon in mijn leven. Ik ben zeer trots en blij om jou naast mijn zijde te hebben en jouw positieve houding in het leven inspireert mij nog steeds dagelijks. Dankjewel voor jouw steun en vertrouwen in mij. Te quiero!

