Prediction Models to Aid Clinical Decision-making in the Emergency Department

Anniek Brink

The research described in this thesis was conducted at the department of Internal Medicine in collaboration with the Emergency Department of the Erasmus University Medical Center, Rotterdam, the Netherlands.

Prediction Models to Aid Clinical Decision-making in the Emergency Department
ISBN: 978-94-6361-641-6

Colophon
Cover-design and lay-out: Optima Grafische Communicatie, Rotterdam, the Netherlands
Cover photo: Maarten Scheffer & Anniek Brink
Print: Optima Grafische Communicatie, Rotterdam, the Netherlands
Printing of this thesis was financially supported by: Pfizer B.V., Chipsoft. None of the above-mentioned organizations could influence or bias the content of this thesis.

Prediction Models to Aid Clinical Decision-making in the Emergency Department

Voorspelmodellen ter ondersteuning van klinische besluitvorming op de spoedeisende hulp

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
woensdag 9 februari 2022 om 15:30 uur

door

Anniek Brink
geboren te Groningen.

Erasmus University Rotterdam

PROMOTIECOMMISSIE:

Promotoren:	prof.dr. R. Zietse prof.dr. S.C.E. Klein Nagelvoort - Schuit prof.dr. H.F. Lingsma
Overige leden:	prof.dr. J.L.C.M. van Saase prof.dr. F.U.S. Mattace Raso prof.dr. P. Nanayakkara
Copromotor:	dr. J. Alsma
Paranimfen:	D.L. Overbeek R. Schuttevaer

"Safari ya ya maili elfu moja huanza na hatua moja"

Swahili proverb

CONTENTS

Chapter 1	General introduction	9
Section I	**Prediction Models for Admission in the Emergency Department**	
Chapter 2	Predicting inhospital admission at the emergency department: a systematic review	25
Chapter 3	Predicting admission in the older population in the emergency department: the cleared tool	45
Section II	**Prediction Models for Mortality in the Emergency Department**	
Chapter 4	Prediction models for mortality in adult patients visiting the emergency department: a systematic review	69
Chapter 5	Predicting mortality in patients with suspected sepsis at the emergency department; a retrospective cohort study comparing qsofa, sirs and national early warning score	93
Chapter 6	Performance of early warning scores predicting 30-day mortality in older emergency department patients	111
Chapter 7	Predicting 30-day mortality using point-of-care testing; an external validation and derivation study	123
Section III	**Drug Prescription and Adherence in the Emergency Department**	
Chapter 8	Non-adherence to antimicrobial guidelines in patients with bloodstream infection visiting the emergency department	137
Chapter 9	The association of body temperature with antibiotic therapy and mortality in patients attending the emergency department with suspected infection	157
Chapter 10	Appropriate empirical antibiotic therapy and mortality: conflicting data explained by residual confounding	175
Chapter 11	Drug non-adherence is a common but often overlooked cause of hypertensive urgency and emergency at the emergency department	191

Section IV **Summary**
Chapter 12 General discussion, future perspectives and conclusions 211
Chapter 13 English summary 233
Chapter 14 Nederlandse samenvatting 241

Appendices References 251
Authors and affiliations 275
Abbreviations 281
Phd portfolio 291
Bibliograph 293
Dankwoord 303
About the author 313

chapter 1

General Introduction

THE EMERGENCY DEPARTMENT

The emergency department (ED) is a facility for patients suffering from accidents or medical emergencies. The ED is the main entrance of the hospital with specialized medical personal where initial diagnostics and emergency treatments are performed[1]. Most general hospitals in the Netherlands have an ED. In 2016 there were 89 EDs in the Netherlands, of which 86 were operational 24 hours a day[2]. The vast majority of these EDs are open access, allowing patients with or without referral to present to the ED and receive proper medical treatment. However, some EDs only admit patients who are referred by a general practitioner, medical specialist or ambulance. The Dutch health-care system strongly encourages to see a general practitioner before visiting the ED to avoid unnecessary ED visits. Self-referral is discouraged in order to limit ED entrance to those patients who do not require emergency care. For these patients it has financial consequences as the health insurance amounts a mandatory excess to them when visiting the ED, which does not apply when visiting the general practitioner[3]. Despite this discouragement, EDs offer emergency care to every patient - even the uninsured[4].

Patients that present at the ED have a wide array of symptoms and diseases[5], and therefore healthcare providers require knowledge about a wide spectrum of disease conditions. Most Dutch EDs have emergency physicians, who are certified after a training program of three years[6]. There is a high degree of heterogeneity in both training and practice of emergency physicians, and emphasis is more often placed on surgical skills. In 2012 a new subspecialty of internal medicine was introduced and since then "acute" internists are increasingly present in EDs, where they treat medical patients - mostly with multiple comorbidities - and complex diseases[7, 8].

Patient numbers and acuity vary throughout the day, and situations occur that the number of patients exceed the number of resources (e.g. physicians, nurses, treatment rooms). Therefore, the acuity of patients is assessed immediately on entering the ED. This is done by triage and aims to prioritise patients according to the urgency of their medical presentation[9]. When resources are scarce, patients with high urgency based on triage are seen before patient with low urgency. In this way the patients that are in most need of treatment are treated first. There are several triage systems used in EDs in the Netherlands, but the most well-known system is the Manchester Triage System (MTS)[10]. More recently, the Netherlands Triage System (NTS) was introduced and validated in EDs[11].

The MTS was developed and introduced in the United Kingdom in 1997. The MTS is a complaint based triage model and categorizes patients' urgency with 52 complaint

code flowcharts. These flowcharts consist of presenting symptoms without making any assumptions about the diagnosis. During triage information necessary to use the flowchart is gathered, and eventually the patient will be classified in an urgency category. These categories are indicators of disease severity and a directive for the time to first assessment by a physician. There are five urgency levels:

1. Immediate; immediate assessment by a physician (this code is most frequently already assigned to patients during transportation to the hospital to allow physicians to be present at arrival at the ED),
2. Very urgent; assessment within ten minutes,
3. Urgent; assessment within one hour,
4. Standard; assessment within two hours, and
5. Non urgent; assessment within four hours (this patient could also be treated by a general practitioner)[12].

After triage, patients proceed to either the waiting area or a treatment room depending on their urgency category and the number of available beds in the ED. Subsequently, physicians will perform a primary assessment of the patient. The most commonly used method is the ABCDE-method[13]. Using this approach a patient's condition is systematically evaluated, diagnosing and treating the most life-threatening issues first[14]. In the majority of cases, the patient undergoes diagnostic testing, i.e. blood tests, ECG and radiological imaging[15]. The decision to either admit or discharge the patient is usually made after the results of diagnostic tests are known. This implies that between the initiation of diagnostic testing and the disposition decision normally more than an hour passes - the average processing time for laboratory results.

CROWDING

ED triage is developed to distribute existing resources and personnel proportionately to the acuity of the patients' illness. When the number of patients outweigh the ED resources this is called crowding. Multiple definitions are used[16-18]. Crowding is compromising patient flow in the ED, and is currently one of the greatest challenges facing modern EDs, with a final result that critical and urgently needed treatment can be delayed[19].

Crowding is a multifactorial problem, but the concept of ED crowding is nicely illustrated by the model of *Asplin et al.*[20]. They describe crowding as an aggregation of factors influencing input, throughput and output (*Figure 1.1*). Input includes conditions that influences the ED demand, such as the number of ED visits - both urgent and of non-urgent - and the distance to, number of and occupation of other EDs in the region. Furthermore,

it is affected by seasonal effects (e.g. the influenza season) and weather conditions. An additional factor is the aging population. As the population ages, the frequency that these people get seriously ill also increases. At the same time with increasing age of the patients, they also become increasingly complex with multiple comorbidities. This will lead to an increased workload, not only in numbers but also in disease severity and complexity.

Throughput refers to all factors in the ED that affect patient flow, such as the total number and availability of ED staff and their experience, their ability and power to make decisions as well as wait times (i.e. time until and duration of the primary and secondary survey/assessment, time until diagnostic tests are performed and their results are retrieved, and the time until ED treatment is initiated)[21]. In summary, it is the time between arrival in the ED and the moment the disposition decision is made and communicated.

Subsequently, output consists of elements relating to an inefficient disposition of ED patients. It is commonly thought that output factors contribute most to crowding and that the inability to move patients to inpatient beds is the most named reason[22]. However, also discharge from the ED to external care facilities, such as nursing homes, is a potential bottle neck. Transportation to other facilities is subject to prioritisation, and non-emergency patient transport by ambulance has lower priority than ambulance transport of ill patients. Also the organization of follow-up care, such as outpatient department appointments and medicine prescriptions, can lead to delayed output times.

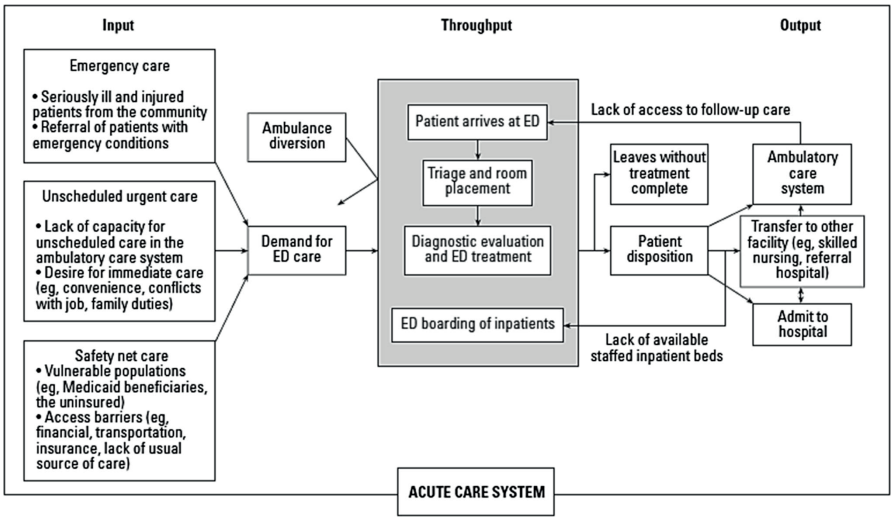


Figure 1.1: Conceptual model of ED crowding.

Crowding has detrimental effects on patients, on the organization within the ED, and beyond the ED. It is associated with poor quality of care, an increase in length of stay (LOS) in the ED, patient dissatisfaction, and last but not least, it causes an increase in morbidity and mortality[23-25]. Crowding also leads to emergency department closure and ambulance blocks (i.e. ambulances are diverted to adjacent hospitals)[26]. This results in a significant delay in time to proper treatment. Crowded EDs not only affects patients, but also have effects on ED staff. Crowding leads to decreased work satisfaction, more absenteeism and less productivity[27, 28]. Beyond the ED, crowding leads to prolonged inhospital LOS, which consequently leads to less inpatient capacity[29].

To combat crowding, multiple measures have been introduced. One of these measures is the introduction of acute medical units (AMU), which are short-stay departments that allow for initial care and observation[30]. Depending on hospital policies, patients can stay up to 72 hours in an AMU, before they are either admitted or discharged. This reduces both the time required to manage inhospital patient capacity and the number of patients admitted towards the wrong ward[31]. Patients who are directly placed in the correct ward undergo less unnecessary movements and subsequently less handovers. Other potential solutions include expanding the hospital capacity and changing admission patterns. In order to change admission patterns, it should be critically evaluated which patients needs to be admitted. By doing so, patients most likely to get admitted can be admitted earlier than when following the current customary pattern.

In the ED, triage systems are used to evaluate acuity and disease severity. The presentation of many medical patients is frequently hard to fit into these systems. Therefore, these systems frequently tend to underestimate (i.e. undertriage) or overestimate (i.e. overtriage) the disease severity. Undertriage leads to potential harm to patients, since time to evaluation and thus time to treatment is longer. Overtriage on the other hand leads to the unnecessary performance of diagnostic tests and early treatment. This can lead to higher expenditure, and even harm for the overtriaged patient, and limits available resources for the undertriaged patient[32, 33].

Especially older patients are at risk of undertriage[34, 35]. This is explained by the fact that older patients present with more non-specific symptoms, and that symptoms may be masked by the presence of multiple co-morbidities[36]. Furthermore, older patients can be frail prior to their ED visit, which makes it difficult to recognize the acute problem for which they visit the ED. Therefore, there is a need to improve the stratification of patients, and prediction models may be a solution.

PREDICTION

Predictions are used to forecast the occurrence of certain events. In medicine it is for example used for prognostic and decision-making purposes[37]. The outcome of interest is most often binary in medical applications (e.g. does the patient get the disease, will the patient survive)[38]. It is either possible to test the performance of a single parameter, i.e. a predictor, or from a combination of parameters, i.e. a prediction model. The performance of prediction models is commonly reflected by two parameters.

First, its discriminative value is of importance. Discrimination is a measure of how well a model can classify patients in those who do and those who do not fulfil the outcome condition. For binary outcomes the area under the curve (AUC) in a Receiver Operating Characteristic (ROC) is a measure of discrimination. This characteristic plots the sensitivity against 1-specificity. When the AUC is 1.0, it means that the model can perfectly discriminate between patients having a high risk for the outcome and the patients with a low risk. On the contrary, when a model has equal chance of assigning a patient with or without the outcome condition to either the high or low risk category, this corresponds to an AUC of 0.5.

The second parameter is calibration. Calibration refers to a statistic to compare the agreement between the expected (i.e. predicted) outcome and the observed outcome in a study population. Ideally, everyone with the predicted outcome also experiences the outcome. Calibration can be assessed using several statistics. The Hosmer-Lemeshow test examines the agreement between observed and expected outcomes with a chi-square distribution. The test should be non-significant in order for the model to be well-calibrated.

Another commonly used measure of calibration is the calibration curve. This is a visualization of the agreement between the observed (y-axis) and expected (x-axis) outcome. Most often we compare the mean predicted outcome to the mean observed outcome, by for example dividing the patients in deciles of predicted outcome. The calibration curve is characterized with a calibration slope, which should ideally be 1 (or on the 45˚ line). The slope is an indicator on how well the predictors within a certain model performs. The intercept in a calibration curve is a measure whether the predictions are systematically too high or too low and is ideally zero[38, 39].

Lack of discrimination and calibration partly explains why prediction models are underused in clinical practice, while several prediction models have been developed for use

in the ED. It would be useful to apply prediction models to predict deterioration in the ED, e.g. an early warning score.

EARLY WARNING SCORES

It is known that adverse events such as intensive care unit (ICU) admission and death are often preceded by clinical signs of deterioration several hours before the event happens[40, 41]. However, both nurses and physicians do not always recognise these signs[42]. This impelled the development and introduction of early warning scores (EWS). The first EWS was published in 1997 as a poster presentation[43], and subsequently, many adjustments have been made[44-48]. These are prediction tools for adverse outcomes such as unexpected ICU admission or death. EWS are mainly based on vital signs, and deviation from the normal values results in increasing points on the EWS. Subsequently the total points are calculated, and depending on the total number of points, action is taken to prevent full deterioration. EWS have been designed for systematic monitoring in hospitalised patients. However, EWS are increasingly used in the ED. In the UK it is recommended to use the national early warning score (NEWS) - which comprises of seven parameters (*Table 1.1*) –in the emergency department to aid the initial assessment of patients, ongoing monitoring and patient triage decisions[44].

Table 1.1: National Early Warning Score

	Score						
Parameter	3	2	1	0	1	2	3
Body temperature in °C	< 35.0		35.1–36.0	36.1–38.0	38.1–39.0	> 39.0	
Heart rate per minute	< 41		41 - 50	51 - 90	91 - 110	111 - 130	> 130
Systolic blood pressure in mmHg	< 91	91 - 100	101 - 110	111 - 219			> 219
Respiratory rate per minute	< 9		9 - 11	12 - 20		21 - 24	> 25
Oxygen saturation in %	< 92	92 - 93	94 - 95	> 96			
Any supplemental oxygen		Yes		No			
Level of consciousness (AVPU)				A			V,P,U

RECOGNITION OF DETERIORATION FOR SPECIFIC DISEASES

Early warning scores are used to recognize deterioration in hospitalized patients. However, patients in the ED differ from hospitalised patients. This is partly due to specific illnesses which are more likely to present at the ED. Amongst these frequently encountered diagnoses are sepsis and hypertensive crisis[49].

Sepsis

An infection can have local or systemic effects, and the severity of the infection can vary from minor to life-threatening. Sepsis is a complex syndrome with high mortality. In sepsis, pro-inflammatory and anti-inflammatory processes take place[50]. A deviation to the pro-inflammatory side may lead to organ dysfunction as a high amount of cytokines induces the attack of target organs. On the contrary, anti-inflammatory overresponse may lead to secondary infections. The clinical presentation of sepsis varies, which makes it hard to detect and treat.

As *Simon Finfer* stated: "Sepsis is like beauty; we know it when we see it, but it is hard to define"[51]. Nevertheless, over the last three decades several groups have tried uniform sepsis definitions. In 1992 the first sepsis definition was introduced. Sepsis was defined as an infection in combination with a systemic inflammatory response syndrome (SIRS)[52]. SIRS consisted of four criteria: a heart rate > 90 beats per minute, a respiratory rate > 20 breaths per minute or $PaCO_2$ of < 32 mmHg, a body temperature < 36 ^{0}C or > 38 ^{0}C, and white blood cell count < 4.0 x 10^9/L or > 12.0 x 10^9/L. The combination of at least two of these criteria and the presence of an infection defined sepsis[52]. Severe sepsis was introduced and defined as sepsis associated with organ dysfunction, hypoperfusion, or hypotension. Septic shock was defined as sepsis with arterial hypotension despite adequate fluid resuscitation.

In 2001 the sepsis criteria were re-evaluated as there was a better understanding of pathophysiology[53]. Nevertheless, the SIRS criteria remained very important for the definition of sepsis. However, additional signs and symptoms of systemic infection, such as elevated C-reactive protein and altered mental status, were included. Furthermore, this consensus meeting led to the introduction of a new staging system of sepsis; the Predisposition, Infection, Response and Organ Failure (PIRO) system. The predisposition refers to premorbid factors such as presence of comorbidities. Next the site, extent, and type of infection largely determine the severity of disease. The response variable refers to the nature and magnitude of the host response. In the end the degree of organ failure is assessed[53].

SIRS criteria however, are not sepsis specific and could also occur without presence of an infection[54]. It has high sensitivity, but low specificity. This led to the development of Sepsis-3 in 2016, in which the SIRS was abandoned[55]. Sepsis was defined as: "life-threatening organ dysfunction caused by a dysregulated host response to infection", and organ dysfunction could be identified by a change in the SOFA score ≥ 2. To facilitate identification of patients potentially at risk of dying from sepsis, the quick Sepsis-related Organ Failure Assessment score (qSOFA) was introduced as a bedside prompt. While the

Sequential Organ Failure Assessment score (SOFA) was developed as a measure of the severity of illness in patients in the ICU, the qSOFA is a simplified version of the SOFA, and is developed for use outside the ICU[56]. The score was simplified using only three parameters; consciousness, blood pressure and respiratory rate (*Table 1.2*).

Table 1.2: qSOFA

qSOFA		
Parameters	Cut-off values	Score
Glasgow Coma Scale	< 15	1
Respiratory rate per minute	≥ 22	1
Systolic blood pressure in mmHg	≤ 100	1

A qSOFA score ≥ 2 is considered positive and suggests a high risk of mortality (≥ 10% chance) in a general hospital population suspected for infection[56]. Patients with a positive qSOFA should be more thoroughly assessed for evidence of organ dysfunction using the SOFA[56].

As sepsis is associated with high mortality rates, timely treatment is paramount[57]. Antibiotic therapy is the most important treatment for sepsis (apart from resuscitation with fluids and oxygen therapy). It is thought that early administration of antibiotic therapy improves patient outcomes. However, at ED entrance it is often unknown which pathogen is causing the sepsis. Results from blood cultures are required in order to confirm sepsis[58]. As blood cultures take at least 24 hours to have conclusive results, physicians therefore have to treat these patients empirically[59].

Hypertensive crisis

A healthy blood pressure is considered to be 120/80 mmHg[60]. When the blood pressure is elevated, it is associated with increased risks of adverse cardiovascular and renal outcomes[61, 62]. In the ED approximately five percent of the patients have an extremely elevated blood pressure, which is commonly defined as a systolic blood pressure ≥ 180 mmHg or diastolic blood pressure ≥ 120 mmHg[63]. This is commonly temporary and a reaction to pain, anxiety or stress[64]. While transient hypertension does not require treatment or follow-up, persistent hypertension does. A distinction should be made between hypertensive emergency and hypertensive urgency. Hypertensive emergency is the situation when acute target organ damage is present, such as acute kidney injury, stroke, and retinopathy. This requires immediate interventions to lower the blood pressure often including intravenously administered antihypertensive treatment and admission to an intensive care unit. When extremely elevated blood pressure is not accompanied with target organ damage, this is called hypertensive urgency. This warrants follow-up

within seven days[65]. While generalisation of EWS developed in general hospital populations to these specific patient groups is questionable, the actual predictive performance of EWS in these groups is unclear.

AIM OF THE THESIS

The aim of this thesis is to develop, validate and improve prediction models for treatment decisions and outcomes in both the general population and specific populations (e.g. elderly, patients with suspected infection) in the emergency department. Specific questions include:

1. Can we reliably predict hospital admission in patients in the ED based on readily available predictors?
2. Can we reliably predict mortality in patients in the ED based on readily available predictors?
3. Can we use prediction tools to predict outcomes in specific patient populations?
4. Does non-adherence to guidelines and treatment have negative effects on patient outcomes?

OUTLINE OF THE THESIS

This thesis is divided in four parts. In the **first part**, prediction tools for admission from the emergency department were investigated. In **chapter 2** I present a systematic overview of current literature regarding prediction models for admission from the emergency department. **Chapter 3** describes the development and validation of an admission model for older ED patients.

The **second part** describes different prediction tools for mortality in emergency department patients. **Chapter 4** includes a systematic review of literature on prediction models for admission in European emergency departments. In **chapter 5** the aim was to externally validate the Sepsis 3.0 criteria for mortality and compare it to the NEWS. **Chapter 6** describes the value of early warning scores based on vital signs, in an older ED population. In **chapter 7** I performed an external validation study of an existing prediction model, including laboratory parameters, for 30-day mortality and tested whether a simplified model yielded comparable performance.

The **third part** consists of four studies on drug prescription and adherence in the ED. The first study in **chapter 8** describes non-adherence of the treating physicians to anti-

biotic guidelines as well as predictors of non-adherence. In **chapter 9** I focussed on the association of body temperature with antibiotic prescription and outcome. **Chapter 10** describes a study on the association between the administration of empirical antibiotic therapy and mortality in patients with a proven bacterial infection. And finally, in **chapter 11** I study the problem of patients' non-adherence to antihypertensive treatment.

The fourth part consists of the general discussion, summaries and appendices. **Chapter 12** contains the general discussion including conclusions and suggestions for future research. **Chapter 13** and **chapter 14** contain the summary of all the studies in English and Dutch respectively.

OVERVIEW OF USED COHORTS

Emergency Department Erasmus University Medical Center cohort

The Emergency Department Erasmus University Medical Center cohort is a retrospective cohort, including all patients who visited the ED of the Erasmus University Medical Center between 2012 and 2017. The database consists of demographic data (age and sex), triage category, mode of arrival to the ED, referral status, vital signs, mortality data, disposition status, laboratory data, and whether radiology tests were performed. Additionally, the ED database was combined with a database from the department of Medical Microbiology and Infectious Diseases. This database contains information on the type of pathogen and their antibiotic susceptibility in all positive blood cultures collected in the ED.

APOP retrospective cohort

The Acutely Presenting Older Patient (APOP) retrospective cohort includes all patients, aged 18 years and older, who visited the ED of the Leiden University Medical Center (LUMC) during the year 2012. Available parameters in this cohort consisted of demographic data (age and sex), triage category, mode of arrival to the ED, type of medical specialist, whether laboratory test were performed and vital signs. The endpoint of hospital admission was available in all patients. Mortality data were only available in patients aged 70-years and older. Only the older ED population, defined as patients aged 70-years and older were used in this thesis.

APOP prospective cohort

The APOP prospective cohort is an observational, multicentre study which took place in two secondary care and two tertiary care hospitals in the Netherlands. Patients were included between September 2014 and January 2017. Patients aged 70 years and older

were included in this study. Within 1 hour of arrival to the ED a battery of tests was performed by trained medical students, such as the 6-CIT score, CAM-ICU and Katz-ADL. Other available parameters were demographic characteristics, mode of arrival to the ED, triage category, vital signs, laboratory test results and geriatric characteristics. Endpoints were 90-day and one year functional decline and mortality.

MST cohort

The Medisch Spectrum Twente (MST) cohort is a single centre retrospective cohort, derived from the National Trauma Registry database. This cohort included all patients aged 70 years and older who presented themselves in the ED during the first three days of every month in 2015. Using this approach we accounted for seasonal variability.

Section I

Prediction Models for Admission in the Emergency Department

chapter 2

Predicting Inhospital Admission at the Emergency Department: a Systematic Review

Anniek Brink
Jelmer Alsma
Lodewijk A.A.M. van Attekum
Wichor M. Bramer
Robert Zietse
Hester F. Lingsma
Stephanie C.E. Schuit

Emerg Med J. 2021 Oct 28;emermed-2020-210902.

ABSTRACT

Background

Emergency department (ED) crowding has potential detrimental consequences for both patient care and staff. Advancing disposition can reduce crowding. This may be achieved by using prediction models for admission. This systematic review aims to present an overview of prediction models for admission at the ED. Furthermore, we aimed to identify the best prediction tool based on its performance, validation, calibration and clinical usability.

Methods

We included observational studies published in Embase.com, Medline Ovid, Cochrane CENTRAL, Web of Science Core Collection, or Google scholar, in which admission models were developed or validated in a general medical population in European EDs including the United Kingdom. We used the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist to assess quality of model development. Model performance was presented as discrimination and calibration. The search was performed on October 11th 2020.

Results

In total 18,539 articles were identified. We included 11 studies, describing 16 different models, comprising the development of nine models, and 12 external validations of 11 models. The risk of bias of the development studies were considered low to medium. Discrimination, as represented by the area under the curve (AUC) ranged from 0.630 to 0.878. Calibration was assessed in seven models and was strong. The best performing models are the models of *Lucke et al.* and *Cameron et al.*. These models combine clinical applicability, by inclusion of readily available parameters, and appropriate discrimination, calibration and validation.

Conclusions

None of the models are yet implemented in EDs. Further research is needed to assess the applicability and implementation of the best performing models in the ED.

Systematic review registration number: PROSPERO CRD42017057975

Keywords

Triage, Crowding, Acute care, Emergency department, Research, Epidemiology

KEY MESSAGES

What is already known on this subject

Several admission prediction tools have been developed with the intention to shorten length of stay at the ED in an attempt to reduce crowding. Implementation of a tool into every day practice has not yet occurred, as this can only be done after validation and calibration in the hospital where it is going to be used. No research to evaluate and compare these models has been published yet.

What this study adds

This systematic review is the first to critically appraise these admission prediction tools. Of the 16 models that we reviewed, only few were adequately developed, validated and calibrated.

INTRODUCTION

It is of great importance to provide timely care for patients in the emergency department (ED). However, sometimes this is compromised by ED crowding, a situation that occurs when there are more patients than available beds in the ED[66]. ED crowding has direct and indirect detrimental consequences for patient care, and ED staff[67, 68]. It leads to an increase in the length of stay (LOS) at the ED, a longer inhospital LOS and an increase in morbidity and mortality[23, 69-71]. There are several causes proposed for the emergence of ED crowding. *Asplin et al.* introduced a conceptual model of ED crowding, visualizing the factors associated with crowding. These factors can be divided into input, throughput and output factors[20]. It is thought that mainly output, i.e. an inadequate disposition of patients, contributes to crowding, which subsequently leads to limited patient flow at the ED. Especially elderly are at risk for a long length of stay and many need to be admitted[72, 73]. Advancing patient disposition may reduce LOS at the ED and thus consequently reduce crowding. The identification of those patients that need admission at ED arrival may help to shorten ED LOS for many patients. Earlier admission (i.e. shorter time in the ED) is associated with improved patient outcomes[74]. Several prediction tools exist to identify patients needing hospital admission. Implementing such a model in clinical practice may alter patient courses and lead to earlier admission[75]. However, a clear overview of literature concerning admission models has not yet been presented. Therefore, the aim of this systematic review is to give an overview of present knowledge on admission prediction models in a general ED population. The secondary aim was to assess the quality of the developed prediction models. As many studies targeted the older population, we will also provide an overview of prediction models developed for this population.

METHODS

The study was conducted and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[76]. We performed a systematic review on prediction models on admission in the ED. The study protocol is registered in the International Prospective Register of Systematic Reviews (PROSPERO) under registration number CRD42017057975.

Eligibility criteria

We aimed to identify all models developed until October 11th 2020 for a non-trauma ED population, or that were applicable to a mixed trauma and non-trauma population. The articles needed to fulfil the following criteria to be considered for inclusion: [1] the pre-

diction tool was developed or validated in an adult population, [2] the prediction model did not have predefined illnesses (e.g. pneumonia) or symptoms (e.g. tachycardia) as inclusion criteria, and [3] the article described a model rather than only individual predictors. Studies that concerned case reports, reviews, or meta-analyses, were excluded. Moreover, the search was restricted to articles written in English. The references of eligible studies were analysed to identify additional articles for inclusion. Because of the heterogeneity between ED systems worldwide, we limited our search to prediction models developed or validated in European EDs including the United Kingdom.

During our initial assessment of the literature we found multiple models for the elderly population. Therefore, we also decided to give an overview of this subgroup of models.

Information sources

The following databases were searched from inception until October 11th 2020: Embase.com, Medline ALL via Ovid, Cochrane CENTRAL Register of Trials via Wiley, Web of Science Core Collection and Google Scholar.

Search

We used among others the following keywords: prediction, risk, hospital admission, emergency department, model, and related synonyms. The queries were developed in Embase.com, and syntax and thesaurus terms were afterwards adjusted for the other databases. The search strategy was created by a biomedical information specialist (WMB). See *Appendix 2.A* for the complete syntaxes.

Study selection

Duplicate articles were removed using Endnote for Windows (Thomson Reuters, version X9) using the method as described by *Bramer et al.*[77]. Two researchers (AB and LA) independently performed the screening of title and abstracts. Conflicting results were discussed in consensus meetings. After screening the abstracts, the full text of the articles was assessed for eligibility by the same researchers and included or excluded in the systematic review. Any remaining disagreement between the first two researchers was discussed with a third investigator (JA).

Data collection process & data items

The following data were extracted from every included article: year of publication, author, journal of publication, country of the study, study period, study design, inclusion and exclusion criteria in their study, study population, hospital setting (i.e. regional hospital, tertiary care hospital), outcome (i.e. number of admissions), model name, parameters within the model, model performance (e.g. discrimination and calibration),

sample size, derivation and/or validation study, calibration method, handling of missing data and patient characteristics (i.e. age, sex). Data were extracted by one investigator (AB) and a random check was performed by a second investigator (JA). This check showed no discrepancies.

Risk of bias in individual studies

The methodological quality and the risk of bias was assessed with the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist, which can be used to describe the reliability, applicability and reproducibility of prediction models[78]. This checklist is applicable for studies deriving a prediction model. The CHARMS checklist assesses risks for bias in the following areas: participant selection, predictor assessment, outcome assessment, model development and analysis. The results were aggregated into a low, moderate or high risk of bias. The risk of bias was assessed by two investigators (AB and LA). Discrepancies were discussed with a third investigator (JA).

Summary measures, data synthesis and analysis

We evaluated the predictive performance of the models, described in terms of discrimination and calibration. Discrimination is a measure of how well a model can distinguish the high-risk from the low-risk group for a certain outcome. It is usually presented as an area under the curve (AUC) in which a value closer to 1.0 indicates better distinction[79].

Calibration reflects the agreement between the expected (i.e. predicted) and observed outcome. This can either be assessed by the Hosmer-Lemeshow goodness of fit test, calibration curve, Wilcoxon signed rank test, Schwarz Bayesian Information Criterion or Brier score[38]. The Hosmer-Lemeshow assesses whether the expected event numbers match the observed event numbers. It provides a chi-square statistic and accessory p-value. A non-significant p-value indicates good calibration. The calibration curve contains a slope and intercept, in which the intercept reflects whether predictions are systematically too low or too high. The calibration slope is a reflection of the predictor effects within the model[38]. The Wilcoxon signed rank test compares two datasets which are not normally distributed. A significant p-value implies that a prediction model performs different in the separate datasets. If a study reported multiple prognostic models or multiple stages of prognostic modelling (e.g. development and validation), data extraction was performed separately for each model or stage. We classified prognostic models as separate models when they included a different set of prognostic variables. Models with identical predictors but for different outcomes were considered validation studies.

Patient characteristics were presented as mean with standard deviation (SD), median with interquartile range (IQR) or numbers with percentages dependent on the distribution of the data. As a result of the heterogeneity of the patient population and the prediction models, a meta-analysis was not possible.

RESULTS

Study selection

In the literature search we detected 18,539 studies, of which 13,017 remained after de-duplication. The exclusion of studies based on title and abstract resulted in 104 full text articles eligible for detailed assessment. The main reasons for excluding articles were that the study described was performed in non-European EDs (n = 29), or that different outcomes were studied (e.g. revisits or length of stay) (n = 29). Finally, we included eleven articles in this systematic review. Full details of study selection are summarized in *Figure 2.1*

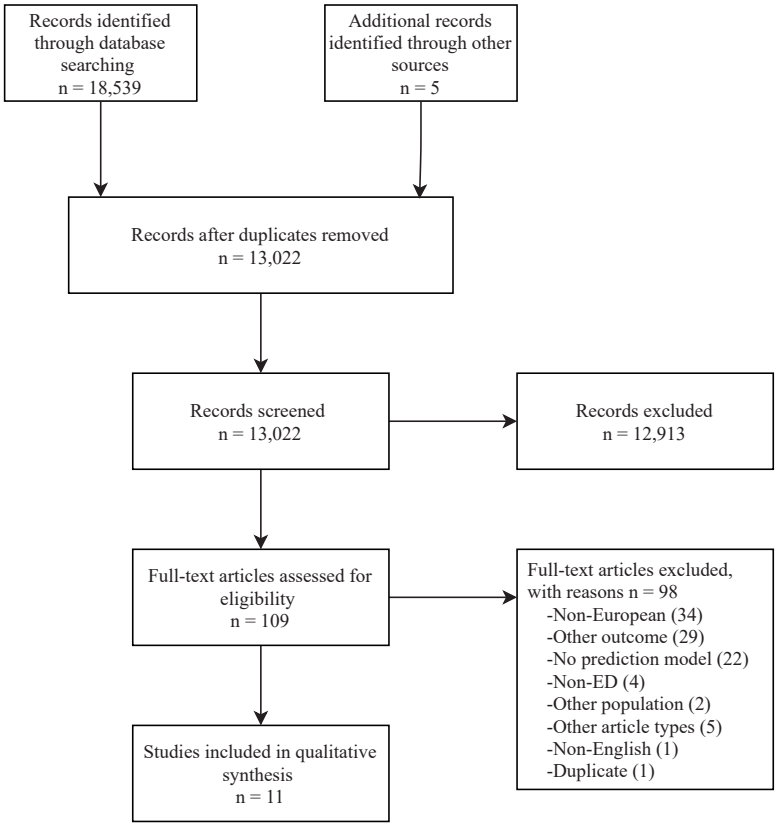


Figure 2.1: Flowchart for literature search

Study characteristics

Study characteristics are summarized in *Table 2.1*. The 11 studies described 16 different models. Two models were tested in two different studies (Identification of Seniors At Risk (ISAR) and Glasgow Admission Prediction Score (GAPS))[38, 80-83]. Most models were constructed with logistic regression. Only three models were developed using machine learning[84, 85].

Seven included studies had a prospective design and the majority of the studies were carried out in a single centre (n = 8). None of the studies assessed prospectively the performance of the model when implemented in day to day practice.

Quality Assessment

The quality of the studies in which a model was developed (n = 5) was assessed using the CHARMS checklist[81, 84, 85, 88, 89]. This considered five studies in which nine models were developed. The results of the CHARMS were aggregated into low, medium and high risk for bias (*Table 2.2*).

The study attrition, referring to the method in which patients were recruited for inclusion, was of good quality in all studies. However, two studies did not describe basic patient characteristics[81, 85]. The outcome was described in all studies. Since the prediction tool had to predict an event in the near future (i.e. admission from the ED), loss to follow up was considered as non-important. Furthermore, the number of patients who were transferred to other hospitals or who left without being seen did not exceed 20%. The number of outcomes in all studies was described and therefore also the number of candidate predictors was satisfying. However, just one study explicitly mentioned that they took into account the number of events per variable to limit overfitting of the model[89]. In general the number of events per variable should at least be ten, meaning that if one hundred events happened the maximum number of predictors in a model is ten.

All studies included parameters that are easily obtainable during triage. Furthermore, one study provided two models which included the triage nurse prediction on admission[84]. This is a subjective parameter and therefore difficult to reproduce. However, in the third model by *Noel et al.* the triage nurse prediction was not included.

The majority of the models (13 / 16) were developed using logistic regression, but in three automated computer techniques were used[84, 85]. All studies used age as a categorical variable in the model. However, it is not clearly described whether categorization of parameters took place before or after inclusion in multivariable analysis.

Table 2.1: Characteristics of studies included for systematic review on prediction models for admission for use in the ED (n = 11)

Study	Population	Country	Publication year	Journal	Study design	Study period	SC / MC	University hospital / Regional hospital	Settings	Study size	Population characteristics
Alam[86]	Adult	The Netherlands	2015	Resuscitation	POCS	7 January 2003 - 15 February 2003	SC	University	ED	274	Mean age (SD): 60 (20) ♂(%): 49.0
Brouns[87]	Older	The Netherlands	2019	BMC Emergency Medicine	ROCS	September 2011 - Augustus 2012	SC	Regional hospital	ED	20,875	Mean age (SD): 77.0 (7.6) ♂(%): 45.8
Cameron'14[81]	Adult	Scotland	2014	Emergency Medicine Journal	RCSS	21 March 2010 - 20 March 2012	MC	University / Regional	ED, AMU, MIU	322,846	NS
Cameron'16[80]	Adult	Scotland	2016	Emergency Medicine Journal	POCS	30 April 2014 - 16 May 2014	SC	University	ED	1,829	Mean age (SD): 47.3 (21.0) ♂(%): 48.4
Di Bari[82]	Older	Italy	2012	Journal of Gerontology	POCS	January - June 2009	SC	Geriatric hospital	ED	1,632	Mean age (SE): 84.0 (5.5) ♂(%): 39.0
Grossmann[35]	Older	Switzerland	2012	Annals of Emergency Medicine	POCS	6 April - 27 April, 2009	SC	University hospital	ED	519	Median age (IQR): 79 (72-84) ♂(%): 45.9
Kraaijvanger[88]	Adult	The Netherlands	2018	Emergency Medicine Journal	POCS	10 - 16 January 2011, 9 - 15 May 2011	MC	University / Regional	ED	3,174	Age: NS ♂(%): 48.8

Table 2.1: Characteristics of studies included for systematic review on prediction models for admission for use in the ED (n = 11) (continued)

Study	Population	Country	Publication year	Journal	Study design	Study period	SC / MC	University hospital / Regional hospital	Settings	Study size	Population characteristics
Lucke[89]	Adult / Older	The Netherlands	2017	Emergency Medicine Journal	ROCS	January - December 2012	SC	University	ED	21,287	Median age (IQR): DV < 70y: 44.8 (28.8 - 57.4) DV ≥ 70y: 78.1 (73.9 - 83.6) VD < 70y: 44.8 (28.4 - 58.0) VD ≥ 70y: 77.9 (73.9 - 83.0) ♂(%): DV < 70y: 54.6 DV ≥ 70y: 47.9 VD < 70y: 54.7 DV ≥ 70y: 50.5
Noel[84]	Adult	France	2018	European Journal of Emergency Medicine	PCSS	8 January - 8 February 2016	MC	University / Regional	ED	9,828	Mean age (SD): DC: 40.8 (22.0), ADM: 61.0 (24.0) ♂(%): DC: 51.1 ADM: 51.0
Salvi[83]	Older	Italy	2012	Rejuvenation research	POCS	January - June 2009	SC	Geriatric hospital	ED	2,057	Mean age (range): 81.7 (65 - 103) ♂(%): 40.0
Zlotnik[85]	Adult	Spain	2016	Computers, Informatics, Nursing	ROCS	January 2011 - December 2012	SC	University	ED	255,668	NS

Abbreviations: ADM, Admitted; AMU, Acute Medical Unit; DC, Discharged; DV, Derivation; ED, emergency department; IQR, Interquartile range; MC, Multicentre; MIU, Minor Injury Unit; NS, Not specified; PCSS, Prospective cross-sectional study; POCS, Prospective observational cohort study; RCSS, Retrospective cross-sectional study; ROCS, Retrospective observational cohort study; SC, Single centre; SD, Standard deviation; VD, Validation; ♂, male.

Table 2.2: Risk of bias in the development studies

	Cameron'14[81]	Kraaijvanger[88]	Lucke[89]	Noel[84]	Zlotnik[85]
Participant selection	Low	Low	Low	Low	Low
Predictor assessment	Low	Low	Low	Moderate	Low
Outcome assessment	Low	Low	Low	Low	Low
Model development	Low	Moderate	Low	Moderate	Moderate
Analysis	Low	Low	Low	Moderate	Low

The risk of bias is evaluated with the CHARMS checklist, which assesses the domains of participant selection, predictor assessment, outcome assessment, model development and analysis. The results are summarized as either low, moderate or high risk of bias.

Description of missing data and handling of missing data was not available for every study[85, 88], one study excluded patients with missing values[84] and two studies compensated missing values[81, 89].

External validation is considered to be the best validation method. Two studies performed external validation[88, 89], whilst two others used internal validation[81, 85]. One study did not perform validation and was therefore considered a high risk of bias[84].

Overall, the models of *Cameron et al.* and *Lucke et al.* were considered to be developed best with an on average low risk on bias in the CHARMS checklist[81, 89].

Participant characteristics

Population size ranged from 274 to 322,846 patients and contribution of male patients ranged from 39.0 to 54.7 percent. Mean age (SD) ranged from 41 (22) to 84 (5.5) years. Four studies included older ED patients, defined as either ≥ 65 years[35, 83, 87] or ≥ 75years[82]. One study compared the older ED population (age ≥ 70 years) with the general adult population[89].

Outcome characteristics

Admission rates varied from 13.6 percent in adults to 59.4 percent in the older patient population.

Variables included in the scoring systems

The number of parameters ranged from one aggregated score (i.e. Emergency Severity Index (ESI)) to thirteen parameters. We subsequently categorized these parameters into demographics, vital signs, interventions, triage, previous care contacts, chief complaint, drug use, mobility and dependency, ED entrance and professional assessment (*Table 2.3*). Most scores included demographic information and triage acuity information as predictors for admission.

Table 2.3: Categorization of parameters in the prediction models

	Model	Demographics	Vital signs	Interventions	Triage	Previous care contacts	Chief complaint	Drug use	Mobility and dependency	ED entrance	Professional assessment
Alam[86]	NEWS		X	X							
Brouns[87]	MTS	X			X						
Cameron'14[81] **& Cameron'16**[80]	GAPS	X	X	X	X	X				X	
Cameron'16[80]	VAS										X
Di Bari[82] **& Salvi**[83]	ISAR	X				X		X	X		
Di Bari[82]	Silver Code	X				X		X			
Grossmann[35]	ESI				X						
Kraaijvanger[88]	Own model	X			X		X			X	
Lucke[89]	Adult model	X	X	X	X	X	X				X
Lucke[89]	Older patient model	X	X	X	X	X	X				X
Noel[84]	TNP										X
Noel[84]	Own model	X			X		X			X	
Noel[84]	TNP + Own model	X			X		X			X	X
Salvi[83]	TRST					X		X	X		X
Zlotnik[85]	Own model LR	X			X		X			X	
Zlotnik[85]	Own model ANN	X			X		X			X	

Abbreviations: ANN, Artificial Neural Network; ED, emergency department; ESI, Emergency Severity Index; GAPS, Glasgow Admission Prediction Score; ISAR, Identification of Seniors At Risk; LR, Logistic Regression; MTS, Manchester Triage System; NEWS , National Early Warning Score; TNP, Triage Nurse Prediction; TRST, Triage Risk Screening Tool; VAS, Visual Analogue Scale. See Appendix 2.B.

General adult population

We included seven studies that developed or validated a model in the general adult population, aged 18 years and over[80, 81, 84-86, 88, 89]. In five studies, in total eight prediction models were developed[81, 84, 85, 88, 89].

Discrimination in the derivation cohorts of these newly developed models ranged from AUC [95% CI] 0.81 [0.790 - 0.820] to 0.878 [0.876 - 0.879]. One study did not provide the derivation AUC, but solely provided the AUCs in the validation population[88]. Four out of five studies also described validation of their developed models. The remaining two studies tested an existing model in their ED. This consisted of the National Early Warning Score[86] and the GAPS[80]. Discrimination in the validation studies ranged from AUC [95% CI] 0.664 [0.599 - 0.728] to 0.876 [0.860 - 0.891]. Calibration was described in five studies. Model characteristics are presented in *Table 2.4*.

Table 2.4: Performance of admission prediction models in the adult population

Study	Model name	Admission, N(%)	Derivation AUC [95% CI]	Calibration method	Calibration derivation	Validation method	Validation AUC [95% CI]	Calibration Validation
Alam[86]	NEWS	130 (47.4)				External	t0: 0.664 [0.599 - 0.728] t1: 0.687 [0.620 - 0.754] t2: 0.697 [0.609 - 0.786]	
Cameron'14[81]	GAPS	NS	0.8778 [0.8764 - 0.8793]	HL GOF test		Split sample	0.8774 [0.8752 - 0.8796]	p = 0.524
Cameron'16[80]	GAPS	745 (40.7)		Wilcoxon Signed Rank test		External	0.876 [0.860 - 0.892]	1.20%
Cameron'16[80]	VAS	745 (40.7)		Wilcoxon Signed Rank test		External	0.875 [0.859 - 0.891]	9.20%
Kraaijvanger[88]	Own model	400 (31.7)	NS	Calibration plot		External	1. 0.88 [0.85 - 0.90], 2. 0.87 [0.85 - 0.89], 3. 0.76 [0.72 - 0.80]	1. α: 0.023, β: 0.974 2. α: 0.05, β: 0.98
Lucke[89]	Own model adults	4,044 (23.6)	0.85 [0.84 - 0.86]	Calibration plot, HL GOF test		External	0.86 [0.85 - 0.87]	p > 0.05
Noel[84]	TNP	2,313 (23.5)	0.815 [0.805 - 0.826]					
Noel[84]	Own model	2,313 (23.5)	0.815 [0.805 - 0.825]					
Noel[84]	TNP + Own model	2,313 (23.5)	0.857 [0.848 - 0.865]					
Zlotnik[85]	Own model LR	34,694 (13.6)	0.8611 [0.8568 - 0.8615]	Calibration plot, HL GOF test	$X^2 = 85.18$	Split sample	0.8568 [0.8508 - 0.8583]	$X^2 = 65.32$
Zlotnik[85]	Own model ANN	34,694 (13.6)	0.8631 [0.8605 - 0.8656]	Calibration plot, HL GOF test	$X^2 = 16.01$	Split sample	0.8575 [0.8540 - 0.8610]	$X^2 = 17.28$

Empty cells mean that specific characteristics were not tested.

Abbreviations: α, calibration intercept; ANN; artificial neural network; AUC, area under the curve; β, calibration slope; CI, confidence interval; GAPS, Glasgow Admission Prediction Score; HL GOF, Hosmer-Lemeshow goodness of fit; LR, logistic regression; N, number; NEWS, National Early Warning Score; NS, not specified; t, timepoint; TNP, triage nurse prediction; VAS, visual analogue scale; X^2, chi-squared

Older ED population

Four studies investigating the older patient population specifically were identified[35, 82, 83, 87]. In these studies, five different models were described, of which three were older patients specific. These models already existed and were used for predicting either frailty or readmission, but not for primary admission. These models included geriatric parameters, such as cognitive impairment and polypharmacy. Discrimination ranged from AUC [95% CI] 0.63 [0.60 - 0.65] to 0.68 [0.66 - 0.70], which represents poor performance. The other two studies investigated triage systems in older patients[35, 87]. The Emergency Severity Index performed best in predicting admission with an AUC [95% CI] of 0.74 [0.73 - 0.75][35]. None of the models were calibrated, nor tested in external populations in these articles.

One study compared the older patient population to the general adult population[89]. This study developed and validated an admission model using temporal validation. The model performed slightly worse in the older ED population, but yielded a good AUC [95% CI] of 0.81 [0.79 - 0.82], which dropped to 0.77 [0.75 - 0.79] after external validation. The positive predictive value and the positive likelihood ratio were higher in the older population. They concluded that further research is needed to investigate the combination of disease severity with frailty to improve prediction of hospital admission in the older patient population. Model characteristics in the older ED population are presented in *Table 2.5*.

Table 2.5: Performance of admission prediction models in the older population

Study	Model name	Admission, N(%)	Derivation AUC [95% CI]	Validation method	Validation AUC [95% CI]	Calibration method	Calibration
Brouns[87]	MTS	4,223 (59.4)		External	0.74 [0.73 - 0.75]		
Di Bari[82]	ISAR	558 (34)		External	0.65 [0.62 - 0.68]		
Di Bari[82]	Silver Code	558 (34)		External	0.63 [0.60 - 0.65]		
Grossmann[35]	ESI	250 (48.8)		External	0.741 [0.734 - 0.747]		
Lucke[89]	Own model older patients	1,817 (43.8)	0.81 [0.79 - 0.82]	External	0.77 [0.75 - 0.79]	Calibration plot, GOF test	p > 0.05
Salvi[83]	ISAR	626 (30)		External	0.68 [0.66 - 0.70]		
Salvi[83]	TRST	626 (30)		External	0.66 [0.64 - 0.69]		

Empty cells mean that specific characteristics were not tested.
Abbreviations: AUC, area under the curve; CI, confidence interval; ESI, emergency severity index; GOF, goodness of fit; ISAR, identification of seniors at risk; MTS, Manchester triage system; TRST, triage risk screening tool

DISCUSSION

The aim of this systematic review was to find and evaluate prediction models for admission used at the ED. We systematically reviewed eleven papers describing the development or validation of sixteen different admission prediction models. Selection of the most appropriate model is based on mainly two qualifications: the model with the lowest probability of overall bias and the highest predictive performances for admission.

Five models reported an AUC over 0.85[80, 81, 84, 85, 89]. The discrimination statistic was highest for the GAPS model[80, 81].

We identified twelve external validations of an admission model. External validation of these models showed substantial variation in performance. This is probably attributable to the fact that some models were tested for a different outcome than they were intended for. Moreover, discrimination may be moderate, because ED populations are heterogeneous. Calibration was executed only in five of the eleven studies. All models reporting calibration were well-calibrated[80, 81, 84, 85, 89].

Apart from the quality of the model, the model should also be easily applicable. In the ED it is useful if the parameters used can be obtained directly, are objective and reproducible. Easily obtainable parameters are predictors that can be retrieved at ED entrance. This will enable immediate use of the prediction model. Several models however use parameters that require (collateral) history which may limit the utility of the prediction model. This information is often not immediately available. The predictors should also be objective, i.e. having a low inter-rater and intra-rater variability. Two studies used the judgment of admission of a healthcare professional as a predictor[80, 84], which is a subjective predictor.

To allow implementation in clinical practice, models should be easy interpretable, or provide applications to enable more complex calculations.

We found that several studies did not report key study details, which made it difficult to judge model utility and makes external validation impossible. With the arrival of machine learning in medical prediction research, models have become more complex. The benefit of machine learning is that models improve from experience. However, machine learning limits insight of how the prediction model works and also limits external validation.

Strengths and limitations

This is to our knowledge the first systematic review on prediction models for admission to the hospital from the emergency department. Strengths of this study include the comprehensive literature search, selection of articles, standard assessment of the articles and the quality assessment using the CHARMS-checklist, which was all performed by two researchers separately. However, also several limitations should be considered. We did limit the inclusion of studies to studies executed in European EDs. We possibly excluded non-European models, which could be applicable to the European ED setting. Even despite only selecting European studies, practice and organization between countries and even between different EDs in the same country are different. Applicability of a prediction tool is dependent on how the healthcare system is organized. Furthermore, the number of included studies might be reduced by only including studies in English. The general limitation of reporting prediction models for admission in the ED is the heterogeneity in ED patients, which is due to epidemiological differences in the populations. This makes it difficult to compare prediction models and to combine these studies in a meta-analysis.

Future directions

None of the studies described implementation, and to our knowledge none of the models are currently implemented in the ED as a prediction tool for admission. The lack of implementation cannot be explained by the discriminative ability which was generally good. Model calibration was lacking in most studies, and therefore it is difficult to judge whether a model, which performs well at group level, is also performing well for individual patients.

Future research should focus on validation, utility of additional predictors, exploration of electronic implementation in patient files to enable the clinical use of prediction models and analysis of their impact. Currently, impact analysis in prediction research is sparse making it difficult to conclude whether a model is worth implementing as an adjunct to clinical evaluation. In the ED it is worthwhile to investigate whether implementation of an admission prediction model reduces ED crowding and improves patient outcomes in terms of a shorter length of stay at the ED and in the hospital.

Furthermore, we recommend that models should be validated and updated to judge generalizability to specific populations prior to implementation. We also recommend that with every external validation study, calibration should be reported.

CONCLUSION

This systematic review identified 16 prognostic models for predicting admission in patients presenting to the ED. The models of *Cameron et al.* and *Lucke et al.* were well developed and have adequate predictive performance. We suggest that the effect of these models on ED LOS and crowding reduction should be examined, given that external validation and potentially updating of the models has taken place for the specific hospital ED.

Appendix 2.A: Search strategy for systematic review of the identification of prediction models for admission in the emergency department

Embase.com
('prediction'/exp OR 'predictive value'/exp OR 'predictive validity'/exp OR 'prediction and forecasting'/de OR 'risk factor'/de OR 'risk assessment'/de OR 'predictor variable'/exp OR (predict* OR ((risk OR hazard) NEAR/3 (factor* OR stratificat* OR assess*))):ab,ti) AND (hospitalization/exp OR 'hospital admission'/exp OR 'hospital readmission'/exp OR 'hospital discharge'/exp OR 'length of stay'/exp OR (hospitalizat* OR hospitalisat* OR rehospitalizat* OR rehospitalisat* OR (hospital NEAR/3 (admis* OR admit* OR readmis* OR readmit* OR discharg* OR stay)) OR ((length OR long* OR short* OR time OR Prolong*) NEAR/3 (stay* OR los))):ab,ti) AND ('emergency care'/exp OR 'emergency patient'/exp OR 'emergency ward'/exp OR 'emergency health service'/exp OR ((emergen* NEAR/3 (ward* OR department* OR patient* OR service* OR admiss* OR admit* OR hospital* OR call*))):ab,ti) AND ('cohort analysis'/exp OR 'follow up'/exp OR 'longitudinal study'/de OR 'retrospective study'/de OR 'prospective study'/de OR 'evaluation study'/de OR model/de OR 'disease model'/de OR 'population model'/de OR 'process model'/de OR simulation/exp OR algorithm/de OR 'validation process'/exp OR 'sensitivity and specificity'/exp OR 'scoring system'/exp OR 'decision tree'/de OR (model OR simulat* OR cohort* OR (follow* NEXT/1 up*) OR followup* OR longitudinal* OR retrospectiv* OR prospectiv* OR evaluation* OR algorithm* OR validat* OR sensitivit* OR specificit* OR score* OR (decision NEXT/1 tree*)):ab,ti) NOT ([Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim) AND [english]/lim NOT ((child/exp OR childhood/exp OR adolescent/exp OR adolescence/exp) NOT (adult/exp OR adulthood/exp)) NOT (pediatrics/exp OR (picu OR nicu OR picus OR nicus OR pediatric* OR paediatric*):ab,ti)

Medline ovid
("Predictive Value of Tests"/ OR "Forecasting"/ OR "Risk Factors"/ OR "risk assessment"/ OR (predict* OR ((risk OR hazard) ADJ3 (factor* OR stratificat* OR assess*))).ab,ti.) AND (exp hospitalization/ OR (hospitalizat* OR hospitalisat* OR rehospitalizat* OR rehospitalisat* OR (hospital ADJ3 (admis* OR admit* OR readmis* OR readmit* OR discharg*)) OR ((length OR long* OR short*) ADJ3 (stay* OR los))).ab,ti.) AND ("Emergency Medical Services"/ OR "emergencies"/ OR exp "Emergency Service, Hospital"/ OR ((emergen* ADJ3 (ward* OR department* OR patient* OR service* OR admiss* OR admit* OR hospital* OR call*))).ab,ti.) AND (exp "cohort studies"/ OR "evaluation study"/ OR exp "Models, Statistical"/ OR "Computer Simulation"/ OR "Models, Theoretical"/ OR Algorithms/ OR "Validation Studies"/ OR exp "sensitivity and specificity"/ OR "Decision Trees"/ OR (model OR simulat* OR cohort* OR (follow* ADJ up*) OR followup* OR longitudinal* OR retrospectiv* OR prospectiv* OR evaluation* OR algorithm* OR validat* OR sensitivit* OR specificit* OR score* OR (decision ADJ tree*)).ab,ti.) NOT (letter OR news OR comment OR editorial OR congresses OR abstracts).pt. AND english.la. NOT ((exp child/ OR exp Infant/ OR adolescent/) NOT (exp adult/)) NOT (exp pediatrics/ OR (picu OR nicu OR picus OR nicus OR pediatric* OR paediatric*).ab,ti.)

Cochrane
((predict* OR ((risk OR hazard) NEAR/3 (factor* OR stratificat* OR assess*))):ab,ti) AND ((hospitalizat* OR hospitalisat* OR rehospitalizat* OR rehospitalisat* OR (hospital NEAR/3 (admis* OR admit* OR readmis* OR readmit* OR discharg*)) OR ((length OR long* OR short*) NEAR/3 (stay* OR los))):ab,ti) AND (((emergen* NEAR/3 (ward* OR department* OR patient* OR service* OR admiss* OR admit* OR hospital* OR call*))):ab,ti) AND ((model OR simulat* OR cohort* OR (follow* NEXT/1 up*) OR followup* OR longitudinal* OR retrospectiv* OR prospectiv* OR evaluation* OR algorithm* OR validat* OR sensitivit* OR specificit* OR score* OR (decision NEXT/1 tree*)):ab,ti)

Google scholar
prediction|"risk|hazard factor|stratification|assessment" hospitalization|rehospitalization|"hospital admission|discharge"|"length of stay" "emergency ward|department|patient|service" model|simulation|cohort|"follow up"|evaluation

Appendix 2.B:
Demographics: Age, sex, marital status, type of residence, insurance status. Vital signs: heart rate, systolic blood pressure, respiratory rate, oxygen saturation, AVPU, body temperature. Interventions: oxygen therapy, phlebotomised blood sample. Triage: ESI category, MTS category, patient acuity score. Previous care contacts: previous ED visit, previous hospital admission, hospitalised past six months, admitted less than one year ago, ED visit in previous thirty days, previous admission and discharge diagnosis. Chief complaint: chief complaint classification, complaint code, visit cause, MTS complaint code. Drug use: six or more medications, polypharmacy, number of drugs in previous three months. Mobility and dependency: difficulties with walking, difficulties with transfer, recent falls, need for help, vision problems, memory problems, more help since hospitalisation. ED entrance: referred by GP, arrived in ambulance, mode of arrival, referral, visit source. Professional assessment: triage nurse prediction, professional recommendation, type of specialist, VAS score.

chapter 3

Predicting Admission in the Older Population in the Emergency Department: the CLEARED Tool

Anniek Brink
Jelmer Alsma
Hans S. Brink
Jelle de Gelder
Jacinta A. Lucke
Simon P. Mooijaart
Robert Zietse
Stephanie C.E. Schuit
Hester F. Lingsma

Neth J Med. 2020 Dec;78(6):357-367.

ABSTRACT

Background

Length of stay (LOS) in the emergency department (ED) is correlated to an extended in-hospital LOS and may even increase 30-day mortality. Older patients represent a growing population in the ED and they are especially at risk of adverse outcomes. Screening tools that adequately predict admission could help reduce waiting times in the ED and reduce time to treatment. We aimed to develop and validate a clinical prediction tool for admission, applicable to the aged patient population in the ED.

Methods

Data from 7,606 ED-visits of patients aged 70 years and older between 2012 and 2014 were used to develop the CLEARED tool. Model performance was assessed with discrimination using logistic regression and calibration. The model was internally validated by bootstrap resampling in Erasmus MC and externally validated at two other hospitals, namely Medisch Spectrum Twente (MST) and Leiden University Medical Centre (LUMC).

Results

CLEARED contains ten predictors: body temperature, heart rate, diastolic blood pressure, systolic blood pressure, oxygen saturation, respiratory rate, referral status, the Manchester Triage System category and the need for laboratory or radiology testing. The internally validated area under the curve (AUC) was 0.766 (95% CI [0.759 - 0.781]). External validation in MST showed an AUC of 0.797 and in LUMC AUC of 0.725.

Conclusions

The developed CLEARED tool reliably predicts admission in elderly patients visiting the ED. It is a promising prompt, although further research is needed to implement the tool and to investigate the benefits in terms of reduction of crowding and LOS in the ED.

Keywords

Aging patient population, Emergency department, Crowding, Prediction model, Triage

INTRODUCTION

Elderly patients represent a growing population in the emergency department (ED)[90-92]. Older patients, defined as aged 70 years and over, are overall more vulnerable than the general adult population. They have less physical endurance, are more likely to have multiple co-morbidities[93] and are also more susceptible to polypharmacy and associated risks[94]. In the ED, elderly patients have a longer length of stay (LOS) compared to younger patients[72, 95-97]. The LOS in the ED is the time between arrival at the ED to the time of discharge from the ED. A prolonged LOS can be the result of delays in triage, consultations and testing (e.g. radiology and laboratory), but also of bed shortage. Elderly patients often have atypical or non-specific presentations of illnesses in the ED[98-100]. Therefore, care for elderly is more complex and as a result elderly patients are more often assessed by multiple specialists, potentially leading to a prolonged LOS[101].

A prolonged LOS in the ED is associated with poorer quality of care and possibly has negative effects on outcomes for the individual patient. Time spent in the ED correlates strongly to the total length of stay in the hospital[102], and periods of longer LOS due to ED crowding are associated with increased inpatient mortality[23].

An earlier decision for admission - instigated by a prediction tool for admission directly after presentation to the ED - might reduce LOS, as it coincides with other waiting times. It has been suggested that new strategies to decrease ED LOS can decrease patient morbidity and healthcare expenditure[103, 104]. Certain patient characteristics, such as aberrant vital parameters, have been shown to be predictive for admission[105, 106], yet it is unknown which set of predictors contributes most to admission.

We aimed to develop and validate a prediction model for admission using non-invasive and readily available variables applicable to the general elderly population at the ED.

MATERIALS AND METHODS

Study design

We performed a retrospective cohort study. Data from one hospital were used for model development and data from two other hospitals were used for external validation.

Setting and participants

For model development, data were acquired from a large consecutive ED cohort in the Erasmus University Medical Center, Rotterdam (Erasmus MC), the Netherlands, including

all patient visits in the ED from 1st of January 2012 until 30th of June 2014. This ED is a level 3 trauma centre and is situated in the largest hospital in the Netherlands, with 30,000 patient visits annually. Elderly patients, defined as people aged seventy years and over, were selected from this database. Both the first visit of a patient as well as repeat visits were included. Patients were excluded when they died on presentation or died during the ED visit.

For external validation, data from the Leiden University Medical Centre (LUMC), which is the academic hospital situated in Leiden, The Netherlands, with approximately 27,000 ED visits annually, and the Medisch Spectrum Twente (MST), a large teaching hospital in Enschede, the Netherlands, with approximately 26,500 ED visits annually, were used. In the LUMC, data were used from an existing database with patients visits in the ED in 2012. For external validation in MST, data were collected from all patients visiting the ED within the first three days of each month in 2015.

Variables and measurement

Outcome was defined as admission or transfer to another hospital for admission and was retracted from the patient records. Basic characteristics including information on sex and age were retrieved from the patient records. Additional ED arrival information was extracted from the patient charts, containing time of arrival and discharge from the ED, triage classification based on the Manchester Triage System[12], vital parameters at arrival (blood pressure (in mmHg), heart rate (per minute), respiratory rate (RR) (per minute), body temperature (in degrees centigrade), peripheral oxygen saturation (SpO_2) (in percentage)), state of consciousness using the AVPU[107] or Glasgow Coma Scale (GCS) scoring system[108] (AVPU/GCS), laboratory testing (yes or no), radiology testing (yes or no) and referral status to the ED (i.e. by ambulance, self-referral, by general practitioner).

After merging all the different variables from the patient record, the patients were coded in order to anonymize the collected data. Only contributors to the study had access to the database. This study was evaluated and approved by the Medical Research Ethics Committee of the Erasmus MC.

Potential predictors were categorized based on to their normal values. Body temperature was categorized in four groups (≤ 35.9, 36.0 - 37.0, 37.1 - 38.4, ≥ 38.5 ˚ C). Heart rate was classified in three categories based on the categories used in the Modified Early Warning Score (MEWS). The MEWS is a guide, developed and validated to aid in recognition of deteriorating patients already admitted in the hospitals. It is based on physiological parameters[109]. In order to facilitate a clear model the original five MEWS categories were reduced to three (≤ 50, 51 - 100, > 100 bpm). Both systolic and diastolic

blood pressure were coded according to the current definition of hypotension (< 90 vs. < 60 mmHg) and hypertension (> 140 vs. > 90 mmHg) for systolic and diastolic blood pressure, respectively. RR was categorized according to the definition of bradypnoea (< 12 times per minute), normopnoea (12 - 20 times per minute) and tachypnoea (> 20 times per minute). SpO_2 was classified in three groups (≤ 92, 93 - 97, ≥ 98%). Finally, the referral status in the ED was coded in three classes based on whether patients were [1] referred by any specialist or general practitioner, [2] or arrived by ambulance, in which case the ambulance nurse decided to present the patient at the ED or [3] were self-referred.

Statistical methods

Univariate logistic regression was used to assess the association between potential predictors and admission. The predictive value was assessed and quantified with the Akaike Information Criterion (AIC) based on the Likelihood Ratio χ^2 ($LR\chi^2$). The AIC is a measure of the relative quality of a model or a parameter and can be used when the database is large and selection on p-value will result in a large number of selected parameters.

We created a missing value category for every parameter. In daily practice, not every parameter is always measured, and with this method the model is still usable with missing parameters. If possible, developing a model with missing values category is statistically sound and a better alternative than excluding missing values, since this will results in losing large amounts of data[78, 110].

We hypothesized that the missing values were more often normal than abnormal, as in clinical practice in patients who appear not severely ill, and therefore have a lower chance of getting admitted, sometimes measurements are not performed or not noted. To test this hypothesis [1] we imputed the missing values with normal values and performed the same analysis as with missing values, and [2] we compared the admission rates in patient groups with missing values to patients without missing values.

The selected predictors from the univariate analysis were combined in a multivariate model and selection of the final set of predictors was based on added values of each predictor (based on the AIC) and clinical knowledge. The performance of the model was calculated using the Area Under the Receiver Operating Characteristic (AUC). During internal validation the AUC was corrected for overfitting using the bootstrapping method[111] on the dataset, with a repetition of the procedure of 500 times.

External validation was performed in the LUMC and MST database. For validation 100 events of 'admission' and at least 100 non-events were required to occur[112]. Based on

admission rates in both LUMC and MST we considered a validation sample of at least 500 patient visits to be sufficient for external validation of our model.

The external validity was examined by calibration and discrimination of the model in the validation samples, using calibration plots and the AUC. In the calibration plot the calculated probability of admission is plotted against the observed admission. The calibration slope is the regression coefficient of the model in which the linear predictor (admission yes or no) is the only parameter. Ideally, the slope is 1[113, 114]. The intercept in the plot indicates whether predictions are systematically too high or too low, and should ideally be zero[115].

In the LUMC database data on whether radiology tests had been performed were not recorded. Therefore, we developed a new model on the data following the same strategy for model development, however leaving out radiology testing. This alternative model was validated in the LUMC sample.

Subsequently, a score chart was developed based on the regression coefficients fitted on the combined data. Therefore, data on radiology testing were imputed for the LUMC database using multiple imputation. An application was built to calculate the chance on admission to facilitate accessibility of the tool in day-to-day practice.

To aid in the decision whether preparations for admission should be started for a specific patient, a specific cut-off point of chance on admission should be determined to guide this decision. Such a cut-off should be based on sensitivity and specificity and the importance of avoiding false-negatives and false positives (i.e. taking action in a patient that in the end does not need to be admitted versus taking no action in a patient that does need to be admitted). Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for a range of possible cut-offs in the combined cohort, MST and LUMC cohort were calculated.

All analyses were undertaken using R statistics version 3.1.3 (2015-03-09)[116]. The foreign library was used to transfer the database from SPSS (version 21) to R[117]. For model development the lrm function of the rms package was used[118]. Finally, the calibration plot was built using the val.prob.ci function, which is an adjustment to the val.prob function of the rms package. For the application we used Rstudio[119].

RESULTS

Participants

The derivation database consisted of 76,663 ED visits between January 2012 and June 2014. Selection on age of 70 and over reduced the number of visits to 7,606 of 5,265 unique patients of whom 54% were admitted to the hospital. In the derivation group 55.8% of the patients were male and the median age was 76 years. The validation dataset consisted of 4,250 patient visits from LUMC of whom 45% were admitted and 563 patient visits from MST, of whom 71% were admitted (*Table 3.1*).

Table 3.1: Patient characteristics in Erasmus MC, LUMC and MST

	Derivation group (n = 7,606)	Validation group MST (n = 653)	Validation group LUMC (n = 4,250)
Parameter			
Age (median ± IQR)	76 (73 - 81)	78 (74 - 83)	78 (74 - 83)
Sex (%)			
Male	4,246 (55.8)	325 (49.8)	2,097 (49.3)
Female	3,360 (44.2)	328 (50.2)	2,153 (50.7)
Temperature (%) (°C)			
≤ 35.9	135 (1.8)	29 (4.4)	243 (5.7)
36.0 - 37.0	3,299 (43.4)	201 (30.8)	1,640 (38.6)
37.1 - 38.4	1,151 (15.1)	144 (22.1)	873 (20.5)
≥ 38.5	484 (6.4)	37 (5.7)	204 (4.8)
Missing	2,537 (33.4)	242 (37.1)	1,290 (30.4)
Heart Rate (%) (bpm)			
< 50	126 (1.7)	14 (2.1)	59 (1.4)
50 - 100	5,039 (66.3)	395 (60.5)	2,734 (64.3)
> 100	1,056 (13.9)	79 (12.1)	593 (14.0)
Missing	1,385 (18.2)	165 (25.3)	864 (20.3)
Systolic Blood Pressure (%) (mmHg)			
< 90	168 (2.2)	8 (1.2)	42 (1.0)
90 - 140	2,730 (35.9)	219 (33.5)	1,478 (34.8)
> 140	3,268 (43.0)	236 (36.1)	1,789 (42.1)
Missing	1,440 (18.9)	190 (29.1)	941 (22.1)
Diastolic Blood Pressure (%) (mmHg)			
< 60	819 (10.8)	64 (9.8)	317 (7.5)
60 - 90	4,041 (53.1)	333 (51.0)	2,395 (56.4)
> 90	1,296 (17.0)	66 (10.1)	598 (14.1)
Missing	1,450 (19.1)	190 (29.1)	940 (22.1)

Table 3.1: Patient characteristics in Erasmus MC, LUMC and MST (*continued*)

	Derivation group (n = 7,606)	Validation group MST (n = 653)	Validation group LUMC (n = 4,250)
Respiratory Rate (%) (x/min)			
< 12	339 (4.5)	0 (0.0)	69 (1.6)
12 - 20	2,091 (27.5)	244 (37.4)	1,612 (37.9)
> 20	254 (16.9)	56 (8.6)	713 (16.8)
Missing	3,893 (51.2)	353 (54.1)	1,856 (43.7)
Oxygen Saturation (%) (%)			
≤ 92	528 (6.9)	64 (9.8)	210 (4.9)
93 - 97	3,130 (41.2)	210 (32.2)	1,323 (31.1)
≥ 98	2,246 (29.5)	207 (31.7)	1,764 (41.5)
Missing	1,702 (22.4)	172 (26.3)	953 (22.5)
Laboratory testing (%)			
Yes	6,217 (81.7)	531 (81.3)	963 (22.7)
No	1,389 (18.3)	122 (18.7)	3,287 (77.3)
Radiology testing (%)			
Yes	4,302 (43.4)	207 (31.7)	NA
No	3,304 (56.6)	446 (68.3)	NA
Arrival (%)			
Self-referral	1,127 (14.8)	30 (4.6)	886 (20.8)
Referred	4,119 (54.2)	501 (76.7)	1,901 (44.7)
Ambulance	2,106 (27.7)	105 (16.1)	1,458 (34.3)
Other	254 (3.3)	17 (2.6)	5 (0.1)
MTS classification (%)			
Immediate/Red	255 (3.4)	22 (3.4)	102 (2.4)
Very urgent/Orange	932 (12.3)	119 (18.2)	1,339 (31.5)
Urgent/Yellow	3,851 (50.6)	410 (62.8)	1,908 (44.9)
Standard/Green	1,592 (20.9)	98 (15.0)	877 (20.6)
Non-urgent/Blue	11 (0.1)	4 (0.6)	17 (0.4)
Missing	965 (12.7)	0 (0.0)	7 (0.2)

Predictors of admission

In the derivation cohort, the strongest predictors of admission were laboratory testing (OR [95% CI]: 13.202 [11.104 - 15.695], p < 0.001) and arrival by ambulance (OR [95% CI]: 5.168 [4.389 - 6.085], p < 0.001) (*Table 3.2*). The categories 'referred' and 'arrival by ambulance' in the predictor referral status had similar odds ratios and were therefore combined. The 'immediate' (red) and 'very urgent' (orange) triage groups in the MTS classification were also combined. These adjustments did not alter the model performance.

Table 3.2: Univariate logistic regression on outcome admission

	Derivation group (n = 7,606)		Validation group MST (n = 653)		Validation group LUMC (n = 4,250)	
Parameter	**Odds Ratio (OR) [95%CI]**	**p-value**	**Odds Ratio (OR) [95%CI]**	**p-value**	**Odds Ratio (OR) [95%CI]**	**p-value**
Age/10	1.127 [1.045 - 1.216]	0.002	1.158 [0.882 - 1.519]	0.290	1.134 [1.103 - 1.248]	0.010
Sex		0.381		0.338		0.004
Female	0.960 [0.877 - 1.051]		0.848 [0.604 - 1.189]		0.839 [0.743 - 0.947]	
Male						
Temperature (°C)		< 0.001		< 0.001		< 0.001
≤ 35.9	1.999 [1.347 - 2.967]		2.015 [0.668 - 6.074]		1.676 [1.274 - 2.205]	
36.0 - 37.0	Ref.		Ref.		Ref.	
37.1 - 38.4	1.339 [1.163 - 1.541]		1.336 [0.791 - 2.254]		1.293 [1.096 - 1.524]	
≥ 38.5	4.942 [3.724 - 6.559]		11.605 [1.550 - 86.867]		8.76 [5.617 - 13.667]	
Missing	0.318 [0.285 - 0.354]		0.400 [0.266 - 0.603]		0.406 [0.347 - 0.475]	
Heart Rate (bpm)		< 0.001		< 0.001		< 0.001
< 50	0.910 [0.637 - 1.300]		0.655 [0.200 - 2.142]		1.258 [0.748 - 2.115]	
50 - 100	Ref.		Ref.		Ref.	
> 100	1.466 [1.274 - 1.687]		4.912 [1.745 - 13.825]		1.743 [1.451 - 2.094]	
Missing	0.229 [0.201 - 0.263]		0.170 [0.115 - 0.253]		0.180 [0.148 - 0.220]	
Systolic Blood Pressure (mmHg)		< 0.001		< 0.001		< 0.001
< 90	1.253 [0.902 - 1.741]		*		7.437 [2.64 - 20.945]	
90 - 140	Ref.		Ref.		Ref.	
> 140	0.931 [0.839 - 1.032]		0.595 [0.366 - 0.967]		0.782 [0.681 - 0.898]	
Missing	0.227 [0.197 - 0.261]		0.127 [0.079 - 0.204]		0.156 [0.127 - 0.190]	

Table 3.2: Univariate logistic regression on outcome admission (*continued*)

	Derivation group (n = 7,606)		Validation group MST (n = 653)		Validation group LUMC (n = 4,250)	
Diastolic Blood Pressure (mmHg)		< 0.001		< 0.001		< 0.001
< 60	1.673 [1.423 - 1.966]		5.122 [1.559 - 16.829]		2.148 [1.668 - 2.765]	
60 - 90	Ref.		Ref.		Ref.	
> 90	1.159 [1.020 - 1.318]		1.027 [0.529 - 1.993]		0.952 [0.796 - 1.139]	
Missing	0.259 [0.226 - 0.295]		0.195 [0.132 - 0.289]		0.184 [0.152 - 0.223]	
Respiratory Rate (x/min)		< 0.001		< 0.001		< 0.001
< 12	1.197 [0.943 - 1.519]		**		0.870 [0.537 - 1.409]	
12 - 20	Ref.		Ref.		Ref.	
> 20	1.459 [1.260 - 1.689]		6.612 [1.557 - 28.082]		2.072 [1.721- 2.495]	
Missing	0.540 [0.485 - 0.602]		0.373 [0.255 - 0.545]		0.419 [0.364 - 0.481]	
Oxygen Saturation (%)		< 0.001		< 0.001		< 0.001
≤ 92	1.823 [1.483 - 2.242]		3.758 [1.429 - 9.882]		3.700 [2.627 - 5.212]	
93 - 97	1.118 [1.001 - 1.248]		1.649 [1.014 - 2.680]		1.127 [0.977 - 1.300]	
≥ 98	Ref.		Ref.		Ref.	
Missing	0.313 [0.274 - 0.358]		0.224 [0.144 - 0.348]		0.228 [0.188 -0.275]	
Laboratory testing		< 0.001		< 0.001		< 0.001
Yes	13.202 [11.104 - 15.695]		10.886 [6.935 - 17.087]		12.138 [9.667 - 15.241]	
No	Ref.		Ref.		Ref.	
Radiology testing		< 0.001		< 0.001		
Yes	2.027 [1.849 - 2.223]		2.414 [1.696 - 3.437]		***	
No	Ref.		Ref.			

Table 3.2: Univariate logistic regression on outcome admission (*continued*)

	Derivation group (n = 7,606)		Validation group MST (n = 653)		Validation group LUMC (n = 4,250)	
Arrival		< 0.001		< 0.001		< 0.001
Self-referral	Ref.		Ref.		Ref.	
Referred	4.559 [3.920 - 5.303]		5.957 [2.664 - 13.322]		1.314 [1.114 - 1.550]	
Ambulance	5.168 [4.389 - 6.085]		9.333 [3.736 - 23.318]		2.204 [1.856 - 2.617]	
Other	3.218 [2.427 - 4.266]		3.333 [0.963 - 11.542]		2.719 [0.452 - 16.359]	
MTS classification		< 0.001		< 0.001		< 0.001
Immediate/Red	1.861 [1.408 - 2.460]		3.898 [0.897 - 16.945]		64.526 [15.870 - 262.350]	
Very urgent/Orange	1.840 [1.575 - 2.149]		4.765 [2.336 - 9.719]		1.958 [1.698 - 2.256]	
Urgent/Yellow	Ref.		Ref.		Ref.	
Standard/Green	0.362 [0.320 - 0.409]		0.236 [0.149 - 0.375]		0.317 [0.262 - 0.383]	
Non-urgent/Blue	0.427 [0.125 - 1.460]		0.130 [0.013 - 1.262]		0.277 [0.079 - 0.965]	
Missing	0.820 [0.711 - 0.944]		****		0.215 [0.026 - 1.790]	

For every parameter p-values are shown. Both OR and 95% CI for every category within a parameter were calculated. *Only eight entries of systolic blood pressure were below 90 mm Hg. Therefore, the categories within this parameter were reduced to three categories (≤ 140, > 140, missing mmHg). **None of the patients at MST had a respiratory rate below 12. ***LUMC did not possess any data on radiology testing. ****MST had no missing values in the MTS classification. Ref.: reference category.

Model development

Based on the AIC we included the following parameters in the final model: laboratory testing, body temperature, heart rate, diastolic blood pressure, systolic blood pressure, SpO_2, respiratory rate, referral status, MTS category and radiology testing, which yielded an AUC of 0.770. Bootstrap resampling decreased the performance by 0.004, resulting in an internally validated AUC of 0.766.

Missing values

Admission rates in patients with a complete set of vital parameters were higher compared to patients with missing values (66% vs. 54 %). We also compared the effect of missing values as a separate category, by replacing them for normal values and by eliminating them. There was no significant effect when either introducing a separate missing category or replacing them by normal values (AUC 0.760 vs. 0.770). However, by taking these values out of the model, the AUC was significantly lower (AUC 0.680 vs. 0.770).

External validation

Patient characteristics in the derivation and validation cohort did not significantly differ, except for admission rates in MST. The univariate effects of the predictors within the validation and derivation cohort were comparable.

However, the OR for MTS classification 'immediate' in the LUMC was higher than in Erasmus MC and MST (OR [95% CI]: 64.526 [15.870 - 262.350] versus 1.861 [1.408 - 2.460] versus 3.898 [0.897 - 16.945].

Discrimination in the LUMC data showed an AUC of 0.725. The calibration plot showed an intercept of -0.308, reflecting the fact that the overall admission rate was lower compared to the development cohort, namely 45%. The calibration slope was 0.826 (*Figure 3.1*).

Discrimination in the MST data showed an AUC of 0.797. However, MST had an admission rate of 71 per 100 patient visits, resulting in a calibration intercept of 1.018 and a calibration slope of 0.904 (*Figure 3.2*).

Score chart

The final model was named 'CalcuLation of the Elderly Admission Risk in the Emergency Department' (CLEARED)-tool and can be used to calculate the probability of admission. We used the parameters from the derivation cohort and introduced a 'hospital factor' to correct for the differences between admission rates of MST (non-academic hospital) and Erasmus MC/LUMC (academic hospital).

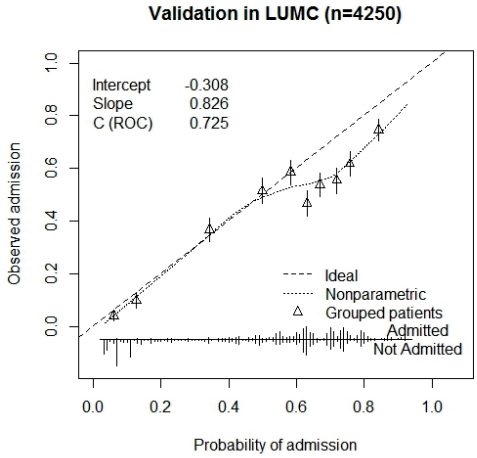


Figure 3.1: Calibration plot LUMC. Comparison of the predicted probabilities and the observed outcome in the LUMC. The diagonal line is the reflection of the ideal situation (predicted probability = observed outcome). The dashed line is the non-parametric relation between the observed and predicted probability. The triangles represent the deciles of the predicted probabilities in the validation set (n = 4,250). The lower part of the figure shows a histogram of the predicted probabilities of admitted and not admitted patients. The intercept, calibration slope and AUC in the validation set is presented.

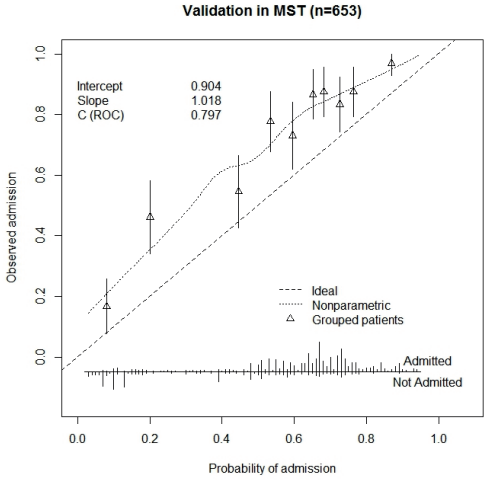


Figure 3.2: Calibration plot MST. Comparison of the predicted probabilities and the observed outcome in the MST. The diagonal line is the reflection of the ideal situation (predicted probability = observed outcome). The dashed line is the non-parametric relation between the observed and predicted probability. The triangles represent the deciles of the predicted probabilities in the validation set (n = 653). The lower part of the figure shows a histogram of the predicted probabilities of admitted and not admitted patients. The intercept, calibration slope and AUC in the validation set is presented.

The formula is presented in *Table 3.3*. the online application is accessible through the following link: http://bit.ly/clearedtool (*Figure 3.3*).

Table 3.3: Regression coefficients for the final model

Parameter	β coefficient	SE	OR
Intercept	-2.3208	0.1072	0.098
Referred			
yes	0.6604	0.0614	1.936
Missing	0.5334	0.1457	1.705
Body temperature (°C)			
≤ 35.9	0.1981	0.1153	1.219
36.0 - 37.0			
37.1 - 38.4	0.1884	0.0561	1.207
≥ 38.5	1.515	0.1250	4.550
Missing	-0.3954	0.0532	0.673
Heart rate (bpm)			
< 50	-0.0670	0.1725	0.935
50 - 100			
> 100	0.0355	0.0632	1.036
Missing	-0.1735	0.1345	0.841
Systolic blood pressure (mmHg)			
< 90	-0.1443	0.1755	0.866
90 - 140			
> 140	-0.0106	0.0486	0.989
Missing	0.4601	0.5922	1.584
Diastolic blood pressure (mmHg)			
< 60	0.5197	0.0787	1.682
60 - 90			
> 90	0.0809	0.0599	1.084
Missing	-0.5494	0.5844	0.577
Respiratory rate (per minute)			
< 12	-0.0880	0.1936	0.916
12 - 20			
> 20	0.1403	0.0623	1.151
Missing	0.0770	0.0499	1.080
Oxygen saturation (%)			
< 92	0.4920	0.0945	1.636
93 - 97	0.1068	0.0470	1.113
≥ 98			
Missing	0.0665	0.0982	1.069

Table 3.3: Regression coefficients for the final model (*continued*)

Parameter	β coefficient	SE	OR
Manchester Triage System			
Immediate / Very urgent	0.4212	0.0534	1.524
Urgent			
Standard	-0.4370	0.0583	0.646
Not-urgent	-0.3087	0.4642	0.734
Missing	0.4022	0.0822	1.495
Laboratory testing			
Yes	1.7004	0.0760	5.476
Radiology testing			
Yes	0.3735	0.0420	1.453
Non-academic hospital	0.9883		2.687

Formula: Admission chance (%) = 1/(1+exp(-(-2.3208 + Referred + Body temperature + Heart rate + Systolic blood pressure + Diastolic blood pressure + Respiratory rate + Oxygen saturation + Manchester Triage System + (Laboratory testing*1.7004) + (Radiology testing*0.3735) + (Non-academic hospital*0.9883))))

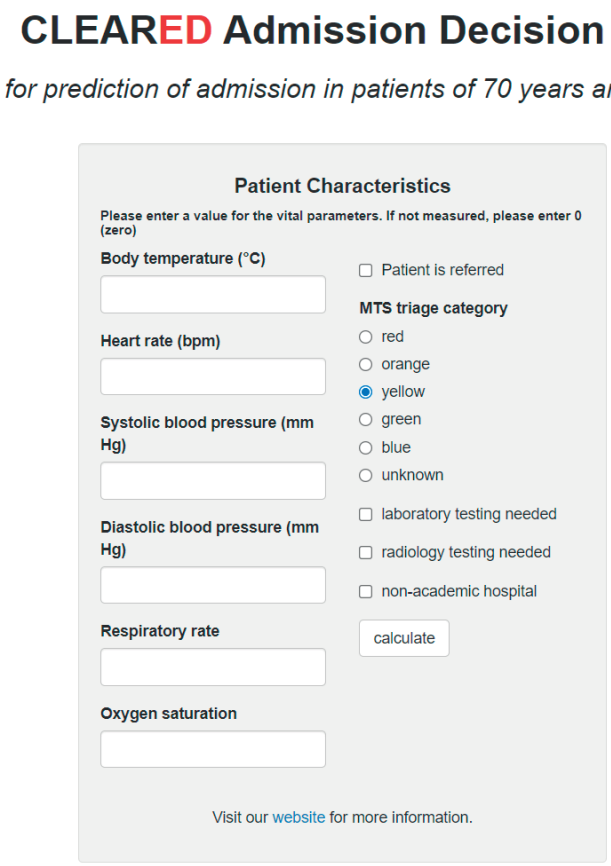


Figure 3.3: The CLEARED tool

We calculated predictive probabilities of the CLEARED tool of admission for decile cut-off points (*Table 3.4*). The predictive probabilities for the separate MST and LUMC cohorts were similar, however PPV was higher for MST. This could be explained by the higher admission rate. The positive predictive value ranged from 0.57 to 0.91. This indicates that in the highest decile 91% of the patients were correctly admitted. An admission cut-off point of 80% would result in the identification of 7.8% (n = 975) who are eligible of earlier admission. Of these patients 86.9% (n = 847) were actually admitted, meaning 13% unnecessary hospital admissions. Patients with a low admission risk (< 80%) had a similar ED LOS than patients who were admitted with a high admission risk (> 80%).

Table 3.4: Predictive probability of the CLEARED tool for different admission probabilities

	All the hospitals combined (%)	95% CI	LUMC	95% CI	MST	95% CI
Admission probability						
60%						
Sensitivity	62.1	60.9 - 63.3	62.0	59.8 - 64.2	61.3	56.7 - 65.8
Specificity	72.4	71.3 - 73.6	70.3	68.4 - 72.2	82.1	75.7 - 87.1
PPV	70.7	69.4 - 71.8	63.2	61.0 - 65.4	89.3	85.3 - 92.4
NPV	64.2	63.0 - 65.3	69.3	67.4 - 71.1	46.6	41.1 - 52.1
Positive tests	45.4	44.5 - 46.3	44.3	42.8 - 45.8	48.7	44.8 - 52.6
70%						
Sensitivity	32.3	31.2 - 33.5	32.8	30.7 - 35.0	29.2	25.1 - 33.6
Specificity	90.1	89.3 - 90.8	90.2	88.9 - 91.4	94.7	90.3 - 97.3
PPV	77.7	76.0 - 79.2	73.4	70.3 - 76.3	93.1	87.3 - 96.5
NPV	55.5	54.5 - 56.5	62.0	60.4 - 63.7	35.4	31.3 - 39.8
Positive tests	21.5	20.8 - 22.2	20.2	19.0 - 21.4	22.2	19.1 - 25.6
80%						
Sensitivity	13.1	12.3 - 14.0	13.1	11.6 - 14.7	13.8	10.9 - 17.4
Specificity	97.9	97.5 - 98.2	98.5	97.9 - 98.9	98.4	95.1 - 99.6
PPV	86.9	84.6 - 88.9	87.8	83.3 - 91.2	95.5	86.6 - 98.8
NPV	51.3	50.4 - 52.3	58.0	56.4 - 59.5	31.9	28.2 - 35.9
Positive tests	7.8	7.3 - 8.3	6.7	6.0 - 7.5	10.3	8.1 - 12.9
90%						
Sensitivity	4.5	4.0 - 5.0	4.3	3.5 - 5.4	2.8	1.6 - 4.9
Specificity	99.5	99.3 - 99.7	99.7	99.4 - 99.9	100.0	97.5 - 100.0
PPV	90.9	87.1 - 93.7	93.3	85.4 - 97.2	100.0	71.7 - 100.0
NPV	49.4	48.5 - 50.3	55.9	54.4 - 57.4	29.7	26.2 - 33.4
Positive tests	2.6	2.3 - 2.8	2.1	1.7 - 2.6	2.0	1.1 - 3.5

Probabilities ranged from 0 to 100%. Sensitivity, specificity, positive- and negative predictive value are presented for different cut-off points for admission. Abbreviations: LUMC, Leiden university medical center; MST, Medisch Spectrum Twente; NPV, negative predictive value; PPV, positive predictive value

However, patients who were eventually discharged despite a positive advice indicated by the CLEARED tool had a significantly longer LOS, median 223 vs. 185 vs. 178 minutes ($p < 0.001$).

DISCUSSION

The aim of this study was to develop and validate a prediction model for admission for the general elderly population presenting to the ED. We successfully developed the CLEARED tool using data from our retrospective cohort study; a prompt that accurately predicts admission of elderly patients visiting the ED. External validation also showed accurate performance. As outlined in the introduction, increased LOS at the ED has detrimental effects on elderly patients. This is not only because the time to adequate treatment is lengthened, but also the result of the stay at the ED itself, which enhances the development of a delirious state[120, 121]. In our hospital after arrival at the ED, initial evaluation and triage of urgency is performed by a nurse followed by a primary survey of the patient by a physician, followed by diagnostics tests. The decision whether or not to admit the patient to the hospital is normally made only after the results of the diagnostics are known. Subsequently, a hospital bed request is made, and the patient awaits transfer to the ward. In this way decision-making about admission is late in the process, which causes a marked increase of LOS at the ED[20]. The use of the CLEARED tool can detect elderly patients, that are to be admitted to the hospital, shortly after their arrival at the ED. From that moment on the admission to the hospital can be organized without delaying the diagnostic and therapeutic process. This can dramatically shorten the LOS at the ED for patients. Further reduction of LOS may be possible when using the CLEARED tool in combination with presence of an acute medical unit.

An increasing number of Dutch hospitals have an acute medical unit. In these hospitals, patients that are identified as in need of admission, can be transferred to such a unit where, after the initial diagnostics and therapeutic interventions are performed at the ED, further diagnostics and treatment can be completed. The incorporation of an acute medical unit in addition to the CLEARED tool can further reduce LOS at the ED which enhances the workflow at the ED and reduces crowding. An acute medical unit is not necessarily a unit, from which every patient gets admitted in-hospital. It can also function as an extension of the ED from which patients can be discharged after a couple of hours of observation. Therefore, the 13% unnecessary admissions can be observed here instead of the ED and thereby optimising patient flow through the ED. This actually reflects current patient care. In our opinion the CLEARED tool should be used in this way to improve patient management.

One of the key points of the CLEARED tool is the inclusion of vital parameters as predictors for admission. Using these parameters we are able to form a better estimate of the severity of illness of a patient, which is the main reason for admission. These parameters are also readily available on arrival at the ED, and are measurements that are routinely performed when patients enter the ED. This makes the CLEARED tool easily applicable.

Another strength of our model is that it is developed using a large database, which reduces the chance of overfitting and limits the uncertainty of the model. Several other prediction models to predict admission have been previously developed. The identification of Seniors at Risk tool (ISAR) is the most well-known screening tool to identify elderly with a high chance of adverse outcomes[122]. This screening tool is composed of six 'yes or no' questions on major topics like cognitive impairment, polypharmacy and previous hospital admission. The main aim of the ISAR tool is to identify patients at risk of loss of functionality after the stay at the hospital and not to identify patients at risk of clinical deterioration and death.

The predictors used in the ISAR tool differ vastly from the predictors in our study. In contrast to the CLEARED tool, the ISAR tool focusses on the pre-existing situation before arriving at the ED where CLEARED focusses on vital parameters at presentation. Therefore, gathering information for the ISAR tool takes more time. The discrimination performance of the ISAR for admission ranges from 0.65 [0.62 - 0.68][82] to 0.68 [0.66 - 0.70][83], which is lower than our model.

Similar to the ISAR is the Triage Risk Screening Tool (TRST), which also comprises of six 'yes or no' questions, such as cognitive and physical impairment, polypharmacy, previous ED visits and hospitalization[123].This model is tested for admission in two studies and performed with an AUC of 0.64 and 0.66, respectively[83, 123]. Other admission prediction tools for elderly are triage based (AUC = 0.73, 0.77, 0.741)[35, 124, 125], the Silver Code (AUC = 0.63)[82] and a tool derived by *Yip et al.* (AUC = 0.713)[126]. Unfortunately, many of these models have not been externally validated.

The higher AUC of our tool suggests that the CLEARED tool is superior in discrimination, although the AUC is, apart from a measure of model performance, also a reflection of the underlying population. Therefore, a prospective validation study should be performed comparing all the existing tools in one large database. A merit of the external validation of the CLEARED tool was that it was performed in both an academic and a large non-academic teaching hospital. The discrimination remained high in both centres, and calibration was well in LUMC, but our model underestimated the chance of admission in MST. This was most likely due to the high admission rate (71%) in this centre and pos-

sibly a difference in population as it is located in different part of the country. A factor was introduced to account for the fact it is a non-academic teaching hospital, although we were unfortunately unable to confirm this with other non-academic hospitals. We expected that patient characteristics might differ between academic and non-academic hospitals; nonetheless it is satisfactory that the CLEARED tool performs well in both settings. This makes it more likely that the model can be generalized to other EDs.

When the CLEARED tool is eventually be implemented in clinical practice, we recommend that implementation takes place in a stepwise process. At first validation of the model should take place to establish its congruency with local protocol. Next, all the caregivers in the ED, especially the triage nurses, should be familiar with the model. To make the model better applicable in clinical practice, we developed an online application, which will automatically calculate the chance of admission after measurement of the parameters. To further determine practical applicability and predictive power a prospective study should be performed.

Our study has several limitations. A general limitation of this study is that the prediction model is developed based on a retrospective database. As a result, not all parameters that are considered in the model are completely accurate, which could have resulted in biased estimates of the effect of certain predictors. For example, it was not assessed whether oxygen therapy was administered before collecting data on both SpO_2 and respiratory rate. Therefore, the predictive value of these variables is probably underestimated in the CLEARED tool. The need for oxygen therapy on its own could also provide to be an independent predictive variable, however prospective study is needed to confirm this.

In addition, respiratory rate was only recorded in just more than half of patient entries. An explanation for this could be that the respiratory rate was only measured when the patient was already in a more severe condition. The predictive value of respiratory rate as an independent variable should be evaluated more in future study.

Inevitably, there were many missing values. In order to deal with missing values and prevent losing data, missing values were categorized in a separate group. Using this method, the model can always be applied, even when not all parameters are recorded. When the CLEARED tool is implemented, possibly vital parameters are more routinely measured, leading to a fall in missing values. When a fall in missing values is observed, the CLEARED tool has to be recalibrated.

In conclusion, the CLEARED tool can accurately predict admission in the elderly. It proved to exceed the performance of comparable tools. However, further research is needed to implement the tool and to evaluate the effect of the CLEARED tool on LOS and crowding in the ED, length of hospitalization and mortality.

Section II

Prediction Models for Mortality in the Emergency Department


chapter
4

Prediction Models for Mortality in Adult Patients Visiting the Emergency Department: a Systematic Review

Anniek Brink
Jelmer Alsma
Aniska W. Fortuin
Wichor M. Bramer
Robert Zietse
Hester F. Lingsma
Stephanie C.E. Schuit

Acute Med. 2019;18(3):171-183.

ABSTRACT

Background

We provide a systematic overview of literature on prediction models for mortality in the emergency department (ED).

Methods

We searched various databases for observational studies in the ED or similar setting describing prediction models for short-term mortality (up to 30 days or in-hospital mortality) in a non-trauma population. We used the CHARMS-checklist for quality assessment.

Results

We found a total of 14,768 articles and included seventeen articles, describing 22 models. Model performance ranged from AUC 0.63 - 0.93. Most articles had a moderate risk of bias in one or more domains.

Conclusions

The full model and PARIS model performed best, but are not yet ready for implementation. There is a need for validation studies to compare multiple prediction models and to evaluate their accuracy.

Keywords

Statistical Models, Mortality, Emergency medicine, Triage, Prognosis

KEY MESSAGES

1. There are many prediction models on mortality in the ED in a non-trauma population that use patient characteristics, vital parameters and results of diagnostic tests.
2. In the models we studied the quality of development varied. All articles we included in our systematic review had some bias on the CHARMS checklist.
3. In general models with more specific, yet difficult to obtain, parameters performed better at the cost of more missing variables in a general population.
4. Sub-optimal model development combined with imperfect model performance prevent implementation of current prediction models in clinical practice at the ED.

INTRODUCTION

Rationale

It is important to provide timely and adequate care for patients in the emergency department (ED). Triage aids physicians in allocating their time and resources. Triage systems, such as the Manchester Triage Score and Emergency Severity Index can identify patients who require earlier treatment, but do not adequately forecast mortality[127]. There is an unmet need for models that objectively determine or forecast which patients have a high risk of mortality. In the case of ED crowding - where there are more patients than treatment rooms, and the waiting room is congested - this is even more important[20]. ED crowding has detrimental consequences for patients resulting in delay in treatment, increased in-hospital length of stay and increased mortality[23, 102].

Development and implementation of a prediction tool for mortality could be helpful to determine which patients benefit most from early treatment, especially during time pressured situations. This can lead to altered treatments regimens, intensified care and prevention of adverse outcomes. Currently, early warning scores (EWS) are used for mortality prediction at the ED, however they were not designed for this purpose[43, 128]. Also, prediction models have been developed and validated for prediction of mortality. Presently, due to the diversity of these models, it is unclear which model is best at predicting mortality in patients presenting at the ED.

Objective

The aim of this systematic review was to give an overview of literature on the most commonly used scoring systems that predict short-term mortality (i.e. up to 30-day or in-hospital mortality) at the ED.

METHODS

Study design

We performed a systematic review on prediction models of short-term mortality in the ED. The systematic review was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[76].The study protocol is registered in the International Prospective Register of Systematic Reviews (PROSPERO) under reference number CRD42017026119.

Eligibility criteria

The search was restricted to studies developing or validating a prediction model at EDs of European hospitals. This was done to minimize the effect of differences in healthcare systems on the EDs, making studies more comparable. Tools developed within a similar setting as the ED, such as an acute medical unit, were also included. Furthermore, the article needed to fulfil the following criteria: [1] the article described a model rather than merely individual predictors, [2] variables within the prediction models were measured at ED presentation, [3] the investigated outcome was short-term mortality (i.e. in-hospital mortality and up to a maximum of 30-day mortality), [4] the study derived or validated a model in a medical (non-trauma) population without selection for specific diseases (e.g. myocardial infarction) or symptoms (e.g. dyspnoea). Studies investigating the association of a triage system with mortality were excluded, as these studies were conducted for another purpose. Where an author published more than one article on the same prediction model, the article describing the model best or the article using the largest sample size was included. Only articles written in English with full-text availability were included.

Information sources

In attempt to identify all relevant studies, the following databases were searched: Embase.com, Medline Ovid, Cochrane CENTRAL, Web of Science Core Collection, and Google scholar. The latest comprehensive search was conducted on the 20th of June 2018.

Search

The search terms for searching the databases were 'prediction models', 'mortality' and 'emergency department' and related synonyms. The queries were developed for Embase.com, and syntax and thesaurus terms were afterwards adjusted for other databases. The search strategy was established by a biomedical information specialist (See *Appendix 4.A* for the complete syntaxes).

Study selection

Articles were deduplicated using EndNote for Windows (Thomas Reuters, version X7). Two investigators (AB and AF) independently reviewed all identified studies for inclusion based on title and abstract. Of the remaining records, full-text was assessed for eligibility by the same investigators. Any discordant results in the selection process were discussed in consensus meetings with a third investigator (JA).

Data collection process & data items

From each included article, the following data were extracted (if available): authors, year and journal of publication, country in which the study was performed, study period, study design, inclusion and exclusion criteria for the patient population, hospital setting (i.e. regional hospital, tertiary care hospital), patient characteristics (i.e. sex, mean age), sample size, the prediction model studied, variable selection of model, time of measuring variables, the outcomes studied, number of outcomes in the investigated population, handling of missing data, model assessment strategy, performance of the prediction model and whether a validation study was executed. If any of this data was missing, it was marked as not specified (NS) in the characteristics table. Data were obtained by two researchers (AB and AF).

Risk of bias in individual studies

The quality of the included articles was assessed according to the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist, which is a guideline that helps to critically evaluate the settings of a study and therefore helps to determine the reliability and applicability of the described prediction models and their outcomes[78]. Only derivation studies were assessed, because they can be evaluated on their predictor selection and model development. Using the CHARMS checklist, we assessed the risks for bias in the following domains: participant selection, predictor assessment, outcome assessment, model development and analysis. Each dimension was assigned a low, moderate or high risk of bias.

Summary measures, data synthesis and analysis

The main outcome was performance of the prediction models. The principle summary measure of model performance was the Area Under the Curve (AUC) in an Receiver Operating Characteristic (ROC) (i.e. how well the model discriminates high-risk from low-risk population). The AUC ranges from 0.5 (no discrimination ability) to 1.0 (perfect discrimination). An AUC > 0.8 is considered to be a reflection of good discrimination[129].

The calibration (i.e. agreement between expected and observed outcomes) within model performance is also an area of interest. Methods to assess calibration were the Hosmer-Lemeshow goodness of fit test[130], Schwarz Bayesian Information Criterion, Brier score or calibration slope. Patient characteristics reported in the articles were presented as mean with standard deviation (SD), median with interquartile range (IQR) or numbers with percentages. If possible, statistics not presented in the articles were calculated from the available data.

RESULTS

Study selection

The electronic literature search identified 14,768 articles. After deduplication 8,099 records remained of which 78 were selected for full-text assessment. Finally, 17 articles were included in the qualitative synthesis of this systematic review (*Figure 4.1*). The latest search was conducted on 20 June 2018.

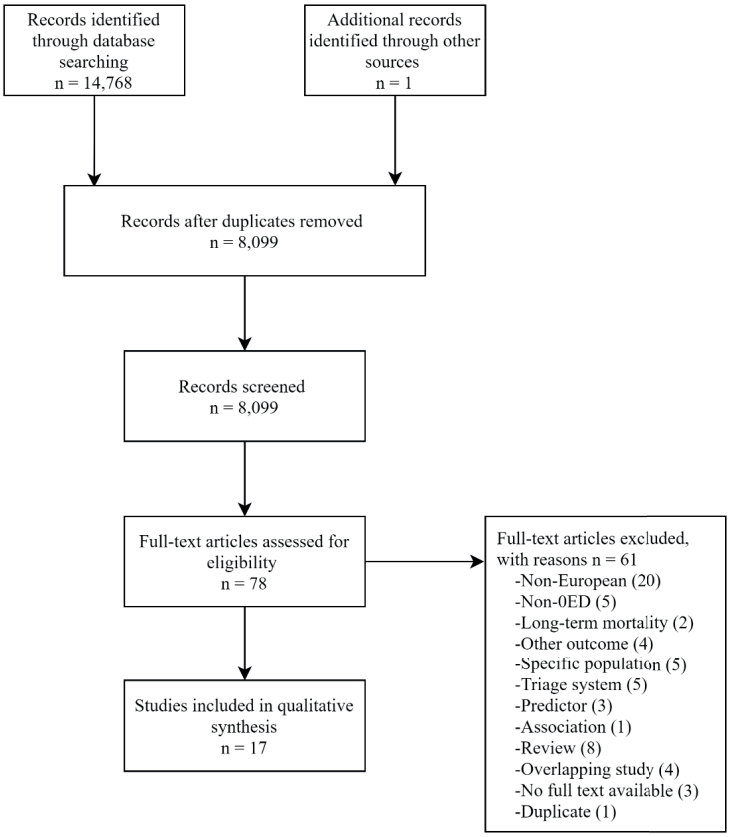


Figure 4.1: Flowchart for literature search

Study characteristics

Seventeen studies investigating 22 different prediction models were included for further analysis (*Table 4.1*). Ten studies focussed on the development or validation of one model, while the remainder developed two or three models. Sample size ranged from 225 to 35,646 patients. Age was either noted as mean (SD), varying from 58.0 to 64.7 years, or as median (IQR), varying from 56 to 71 years. The ratio of male/female was similar in all

articles, with percentage of male ranging from 46.1% to 57.7%. Mortality rates in the study population varied between 0.6% (40/6,947) and 12.7% (711/5,583). Missing data were not always reported, and neither was handling of missing data. Fifteen studies excluded cases with missing values or considered missing values to be normal. Only two studies used imputation techniques to replace missing values[131, 132].

Quality assessment

The quality of studies and susceptibility of bias between studies were assessed using CHARMS. Three studies did not extensively describe patient selection and therefore were considered as having moderate risk of bias[128, 133].

The Simple Clinical Score (SCS) and the Hypotension, Oxygen saturation, low Temperature, ECG changes and Loss of independence (HOTEL) score used subjective predictors (e.g. breathlessness), which can be difficult to reproduce. Therefore, the articles reporting these scores were assigned a moderate risk of bias in predictor assessment[134-136]. Studies that used predictors (e.g. laboratory values) that are not immediately available when a patient presents to the ED were also assigned a moderate risk for bias[133, 137-140].

In the appraisal of outcome two domains were assessed; description of outcome and description of loss of follow-up. Outcome was reported in all articles and resulted in low risk of bias. However, loss of follow-up was described in only five articles, resulting in a moderate risk of bias for the articles that did not provide these data[86, 131, 132, 134, 141]. *Brabrand et al.*[131] and *Coslovsky et al.*[132] described the development process of their models best. The study by *O'Sullivan et al.* did not describe the development process and was therefore considered having a high risk of bias[139]. Most articles entered all the variables with a strong predictive ability in a multivariate logistic regression analysis, however three studies used a backward stepwise regression procedure to identify the best prediction model[131, 132, 142].

Studies including continuous variables in the prediction tool were less likely to have bias[131, 133, 140]. The majority of studies excluded patients with missing data and were therefore assigned a moderate risk of bias. Two studies used imputation methods to replace missing values[131, 132]. Two other studies replaced missing values by normal values[128, 143].

Finally, the quality of analysis used in the articles was assessed. Seven articles did not provide a validation study and were therefore scored a high risk of bias[86, 128, 137, 139, 142, 144, 145].

Table 4.1: Study characteristics

Study	Country	Publication year	Study design	Study period	University hospital / Regional hospital	Settings	Study size, N	Population characteristics
Alam[86]	The Netherlands	2015	prospective observational cohort study	7th January 2013 - 15 February 2013	University hospital	ED	274	Mean age (SD): 60 (20) ♂ (%): 49.0
Brabrand[131]	Denmark	2015	prospective observational cohort study	October 2008 - February 2009, February 2010 - May 2010, March 2011 - July 2011	University hospital / Regional hospital	MAU	1. 2608 2. 2463 3. 2210	Mean age (SD): 1. 62.4 (19.2) 2. 61.1 (19.4) 3. 63.0 (20.8) ♂ (%): 1. 52.1 2. 47.7 3. 46.1
Bulut[142]	Turkey	2014	prospective observational cohort study	October 2011 - April 2012	University hospital	ED	2000	Mean age (SD): 61.41 (18.92) ♂ (%): 52.0
Coslovsky[132]	Switzerland	2015	prospective observational cohort study	October 2009 - October 2010	University hospital	ED, ICU	8606	Mean age (SD): 58 (20) ♂ (%): 62.0
Cournane[137]	Ireland	2017	retrospective observational cohort study	2002 - 2016	Secondary care centre (large central teaching hospital)/regional hospital	ED, AMUA	50612	Median age (IQR): 62.1 (40.3 - 78.4) ♂ (%): 48.7
Duckitt[141]	UK	2007	prospective observational cohort study	July 2003 - November 2003 , October 2005 - November 2005	Regional hospital (general)	EAU	1. 3184 2. 1102	Mean age (range): 1. 71 (17 - 106), †83 (32 - 99) 2. 73 (19 - 102). †81 (43 - 98) ♂ (%): 1. 48.0, † 46.0 2. 45.0, † 51.0

Table 4.1: Study characteristics (*continued*)

Study	Country	Publication year	Study design	Study period	University hospital / Regional hospital	Settings	Study size, N	Population characteristics
Goodacre[144]	UK	2006	retrospective observational cohort study	July 1996 - January 2001, sampled	University hospital	ED	5583	Mean age: 63.4 ♂ (%): 57.7
Groarke[128]	Ireland	2008	prospective observational cohort study	30 day period; 8 - 19.00	Regional hospital (general)	MAU	225	Mean age (SD): 64.7 (19.1) ♂ (%): 51.6
Kellett (2006)[134]	Ireland	2006	prospective observational cohort study	17 February 2000 - 29 January 2004	Regional hospital	AMU	1. 6736 2. 3228	Mean age (SD): 1. 61.9 (20.3) 2. 62.1(20.2) ♂ (%): 1. 52.5 2. 50.1
Kellett (2008)[135]	Ireland	2008	prospective observational cohort study	17 February 2000 - 29 January 2004	Regional hospital	AMU	1. 6947 2. 3343	Mean age (SD): 1. 61.6 (20.4) 2. 61.9 (20.3) ♂ (%): 1. 52.7 2. 50.4
Kristensen[138]	Denmark	2017	prospective observational cohort study	22 September 2009 - 28 February 2010, 4 September 2013 - 13 December 2013	University hospital	ED	1. 5371 2. 5738	Median age (IQR) 1. 63.8 (46.92-76.52) 2. 63.0 (46.0-76.0) ♂ (%): 1. 48.0 2. 49.4
Merz[145]	Bern	2011	prospective observational cohort study	11 June 2007 - 11 January 2008	University hospital	ED	4388	Median age (IQR): 61.0 (44.3 - 74.1) ♂ (%): NS

Table 4.1: Study characteristics (*continued*)

Study	Country	Publication year	Study design	Study period	University hospital / Regional hospital	Settings	Study size, N	Population characteristics
Olsson[143]	Sweden	2004	prospective observational cohort study	October 1995 - November 1996	University hospital	ED	11751	Mean age (SD): 61.9 (20.7) ♂ (%): 48.4
O'Sullivan[139]	Ireland	2012	retrospective observational cohort study	1 January 2005 - 31 December 2010	Secondary care centre (large central teaching hospital)	ED, AMU	20848	Median age (IQR): 56.5 (37.2 - 75.8) ♂ (%): 48.1
Silke[133]	Ireland	2010	prospective observational cohort study	1 January 2002 - 31 December 2007, 17 february 2000 - 29 January 2004	Secondary care centre (large central teaching hospital)/regional hospital	ED, AMAU	10712, 13182, 3597	Median age (IQR): 58.9 (37.9 - 75.6) ♂ (%): 48.1
Slagman[140]	Germany	2015	retrospective observational cohort study	15 February 2009 - 14 February 2010, 15 February 2011 - 14 February 2012	University hospital	ED	34333, 35646	Median age (IQR): 56 (38 - 71),†/ 71 (63 - 81) ♂ (%): 48.8
Straede[136]	Denmark	2014	prospective observational cohort study	2 October 2008 - 19 February 2009	Regional teaching hospital	MAU	1080, 1470	Mean age (SD): 62.4 (19.2), ♂ (%): 52.1

Characteristics of studies included for systematic review on prediction models for mortality for use in the ED (n = 15). Abbreviations: AMAU, Acute Medical Assessment Unit; AMU, Acute Medical Unit; EAU, Emergency Assessment Unit; ED, emergency department; ICU, Intensive Care Unit; IQR, Interquartile range; MAU, Medical Assessment Unit; NS, Not specified; SD, Standard deviation; ♂, male; †, non-survivor.

Five studies used internal validation[132, 134, 135, 141, 143], resulting in a moderate risk. Four studies performed external validation and were scored as having a low risk of bias[131-133, 136, 138] (*Table 4.2*). Overall, none of the models scored a low risk of bias on all individual domains. Seven studies had a high risk on one of the domains within CHARMS[86, 128, 137, 139, 142, 144, 145]. The models by *Coslovsky* and *Brabrand* scored best with an overall low risk of bias on all domains[131, 132].

Table 4.2: Risk of bias in the development studies

	Participant selection	Predictor assessment	Outcome assessment	Model development	Analysis
Alam (2015)[86]	Low	Low	Low	Moderate	High
Brabrand (2015)[131]	Low	Low	Low	Low	Low
Bulut (2014)[142]	Low	Low	Moderate	Moderate	High
Coslovsky (2015)[132]	Low	Low	Low	Low	Low
Cournane (2017)[137]	Moderate	Low	Moderate	Moderate	Moderate
Duckitt (2007)[141]	Low	Low	Low	Moderate	Moderate
Goodacre (2006)[144]	Low	Low	Moderate	Moderate	High
Groarke (2008)[128]	Moderate	Moderate	Moderate	Moderate	High
Kellett (2006)[134]	Low	Low	Low	Moderate	Moderate
Kellett(2008)[135]	Low	Low	Moderate	Moderate	Moderate
Kristensen (2017)[138]	Low	Low	Moderate	Moderate	Low
Merz (2011)[145]	Low	Low	Moderate	Low	High
Olsson (2004)[143]	Low	Low	Moderate	Moderate	Moderate
O'Sullivan (2012)[139]	Low	Low	Moderate	Moderate	High
Silke (2010)[133]	Moderate	Moderate	Moderate	Moderate	Low
Slagman (2015)[140]	Low	Moderate	Moderate	Moderate	Moderate

The risk of bias is assessed by the CHARMS checklist, which assesses the domains of participant selection, predictor assessment, outcome assessment, model development and analysis. The results are summarized as either low, moderate or high risk of bias.

Variables included in the scoring systems

Median number of included predictors was 6 (IQR 5 - 8.5). Most prediction tools were primarily based on vital parameters (e.g. heart rate, oxygen saturation and body temperature). Eight models included laboratory results. Three prediction tools were solely based on biomarkers combined with patients' age and sex[138-140]. Predictors were categorized in patient characteristics, ED presentation, vital parameters, laboratory values, interventions and tests (*Table 4.3*).

Table 4.3: Variables within the prognostic models

Variables	NEWS[86]	Full model Brabrand[131]	PARIS[131]	MEWS[142]	REMS[142-144]	Full model Goslovsky[132]	Nurse risk estimate model[132]	Worthing PSS[141]	EWS[128,141]	RAPS[143,144]	SCS[134,136]	HOTEL[135]	VSS[145]	aAISS[139]	aAISS +comorbidity[139]	MARS[133]	MARS lab- only[133]	Full model Slagman[140]	EPICS[140]	Full model Kristensen[138]	Full model Courtane[137]
Patient characteristics																					
Age		X	X		X	X					X			X	X	X	X	X	X	X	X
Los of independence		X	X									X								X	
Diabetes											X										
Unable to stand unaided, or a nursing home resident											X										
Prior to current illness, spent some part of daytime in bed											X										
Comorbidities															X						
ED Presentation																					
APACHE II diagnostic category						X															
Breathless on presentation											X										
New stroke on presentation											X										
Coma without intoxication or overdose											X										
Altered mental status without coma, intoxication or overdose, and aged > 50											X										
Seizures													X								
Previous ED visit 12 months						X															
Vital parameters																					X
Body temperature	X			X				X	X		X	X				X					
Heart rate	X			X	X			X	X	X	X		X			X					
Systolic blood pressure	X	X	X	X				X	X	X	X	X	X								
MAP					X	X										X					
Capillary refill time						X															
Respiration rate	X	X	X	X	X			X	X	X	X		X			X					
Oxygen saturation	X				X			X	X		X	X	X								
SaO2/FiO2		X	X			X															
AVPU	X			X				X	X												
GCS					X	X				X			X								
Laboratory values																					
Albumin														X	X						
Troponin Positive														X	X						
Urea														X	X	X	X			X	X

Table 4.3: Variables within the prognostic models (*continued*)

Variables	NEWS[86]	Full model Brabrand[131]	PARIS[131]	MEWS[142]	REMS[142-144]	Full model Goslovsky[132]	Nurse risk estimate model[132]	Worthing PSS[141]	EWS[128, 141]	RAPS[143, 144]	SCS[134, 136]	HOTEL[135]	VSS[145]	aAISS[139]	aAISS +comorbidity[139]	MARS[133]	MARS lab- only[133]	Full model Slagman[140]	EPICS[140]	Full model Kristensen[138]	Full model Cournane[137]
RDW														X	X			X	X		
WBC														X	X	X	X	X			X
Sodium														X	X		X	X			X
Potassium														X	X	X	X	X		X	
Haematocrit																X	X			X	X
CRP																		X	X	X	
Creatinine																		X		X	
Platelets																		X			
Sodium																				X	
RBC																		X		X	
Interventions & tests																					
Mechanical ventilation						X							X							X	
Supplemental oxygen	X																				
Abnormal ECG											X	X									
Other																					
Nurse risk estimate							X														

Abbreviations: APACHE II, Acute Physiologic Assessment and Chronic Health Evaluation II; AVPU, Alert, Verbal, Pain, Unresponsive; CRP, C-reactive protein; ECG, Electrocardiography; ED, emergency department; GCS, Glasgow Coma Scale; MAP, Mean arterial pressure; RBC, Red blood cell count; RDW, Red cell distribution width; SaO2/FiO2, pulse oximetry saturation / fraction of inspired oxygen; WBC, White blood cell count; NEWS, National early warning score; EWS, Early warning score; MEWS, Modified early warning score; SCS, Simple clinical score; HOTEL, Hypotension, oxygen saturation, low temperature, ecg changes, loss of independence; PARIS, Blood pressure, age, respiratory rate, loss of independence, peripheral oxygen saturation; REMS, Rapid emergency medicine score; RAPS, Rapid acute physiology score; PSS, Physiological scoring system; VSS, Vital sign score; AISS, Acute illness severity score; EPICS, Emergency processes in clinical structures; MARS, Medical admissions risk system; MTS, Manchester Triage System.

Outcomes

Outcomes were defined as mortality up to 30 days or in-hospital mortality. Further distinction was made in 24-hour mortality[134-136], five-day mortality[133], seven-day mortality[131, 133], and 30-day mortality[86, 134, 136-139].

Model performance

Discrimination was described in all studies, except by *Groarke et. al*[128]. Based on the reported sensitivity and specificity we approximated the AUC for this study. Eleven models provided an AUC < 0.8, of which five showed a poor discriminative ability (MEWS AUC =

0.630, EWS AUC = 0.68 / 0.656, RAPS AUC = 0.64 / 0.652)[128, 141-144]. The MARS model had the best discriminative ability (AUC = 0.93, 95% CI [0.92 - 0.94][133].

Calibration was measured for eleven models, of which eight used the Hosmer-Lemeshow goodness of fit test[131, 133, 135, 136, 141, 143]. One article combined the calibration by the Hosmer-Lemeshow goodness of fit test with the Schwarz Bayesian Information Criterion[135], and another article developed a calibration curve and reported the calibration slope and calculated the Brier score[132]. In two studies the Hosmer-Lemeshow goodness of fit test yielded a significant p-value in the derivation, which proves bad calibration[133, 143](*Table 4.4* and *Figure 4.2. a-e*).

Validation

Six studies performed an internal validation analysis, divided in temporal validation[141], split-sample validation[132, 134, 135, 143], cross-validation[138] and bootstrap resampling validation[132]. External validation analysis was performed in five studies[131, 133, 136, 138, 140]. The AUC for all external validation studies was high, ranging from 0.837 and 0.960. Calibration within the validation studies was performed for twelve models, and only the PARIS model scored poorly in one validation dataset[131](*Table 4.5*).

Table 4.4: Performance of the developed prediction models

Study	Prediction model	Mortality, N(%)	Discrimination, AUC [95% CI] /(±SE)	Calibration method	Calibration	Other performance measurements
24-hour mortality						
Kellett (2008)[135]	HOTEL	40 (0.6) , 19 (0.6)	0.865 [0.793 - 0.937]	Schwarz BIC, HL GOF test	422.89, χ^2 = 1.49 (0.83)	NS
Kellett (2006)[134]	SCS	40 (0.6), 19 (0.6)	0.902 (±0.019)	NS	NS	NS
5-day mortality						
Silke[133]	MARS	648 (6.05) , 171 (4.75)	0.93 [0.92- 0.94]	HL GOF test	χ^2 = 5.66 (0.315)	NS
Silke[133]	MARS lab only	648 (6.05)	0.90 [0.89 - 0.90]	HL GOF test	χ^2 = 11.65 (0.167)	NS
7-day mortality						
Brabrand[131]	Full model	76 (2.5)	0.87 [0.82 - 0.93]	HL GOF test	P = 0.97	NS
Brabrand[131]	PARIS	76 (2.5)	0.86 [0.80 - 0.91]	HL GOF test	P = 0.42	NS
Silke[133]	MARS	788 (5.98)	0.91 [0.90 - 0.93]	HL GOF test	χ^2 = 17.98 (0.02)	NS
30-day mortality						
Alam[86]	NEWS	11 (4.0)	0.768 [0.618 - 0.919]	NS	NS	NS
Kellett (2006)[134]	SCS	316 (4.7)	0.858 (±0.009)	NS	NS	NS
O'Sullivan[139]	aAISS	(4.8)	0.90 [0.89 - 0.90]	NS	NS	NS
O'Sullivan[139]	aAISS + comorbidity	(4.8)	0.89 [0.88 - 0.89]	NS	NS	NS
Kristensen[138]	Full model	284 (5.3)	0.886 [0.861 - 0.911]	BS	BS 4.11 (3.54 - 4.70)	NS
Cournane[137]	Full model	(4.6-7.0)	0.85 [0.85 - 0.86]	NS	NS	NS
Inhospital mortality						
Bulut[142]	MEWS	153 (7.65)	0.630 [0.608 - 0.727]	NS	NS	NS
Bulut[142]	REMS	153 (7.65)	0.707 [0.686 - 0.727]	NS	NS	NS
Coslovsky[132]	Full model	398 (4.6)	0.922 [0.916 - 0.927]	BS , CS	BS 0.028, CS 0.95	NS
Coslovsky[132]	Nurse risk estimate model	398 (4.6)	0.78	BS	BS 0.040	NS

Table 4.4: Performance of the developed prediction models (*continued*)

Study	Prediction model	Mortality, N(%)	Discrimination, AUC [95% CI] /(±SE)	Calibration method	Calibration	Other performance measurements
Duckitt[141]	Worthing PSS	270 (8.0)	0.74 [0.71 - 0.77]	HL GOF test	p = 0.119	NS
Duckitt[141]	EWS	270 (8.0)	0.68 [0.65 - 0.71]	NS	NS	NS
Goodacre[144]	RAPS	711 (12.7)	0.64 [0.59 - 0.69]	NS	NS	NS
Goodacre[144]	REMS	711 (12.7)	0.74 [0.70 - 0.78]	NS	NS	NS
Goodacre[144]	Full model	711 (12.7)	0.81 [0.78 - 0.84]	NS	NS	NS
Groarke[128]	EWS	8 (3.6)	0.656 *	NS	NS	OR 2.19 [1.41 - 3.39]
Merz[145]	VSS	316 (7.2)	0.72 [0.53-0.91]	NS	NS	NS
Olsson[143]	RAPS	285 (2.4)	0.652 (± 0.019)	NS	NS	NS
Olsson[143]	REMS	285 (2.4)	0.852 (± 0.014)	HL GOF test	$\chi^2 = 62$ (< 0.0001)	NS
Slagman[140]	Full model	634 (1.8)	0.863 [0.848 - 0.877]	NS	NS	NS
Slagman[140]	EPICS	634 (1.8)	0.866 [0.853 - 0.878]	NS	NS	NS

Performance of the developed prediction tools for mortality, divided by the time-frame of mortality**.** Abbreviations: AISS, Acute illness severity score; AUC, Area under the curve; BS, Brier score; CI, Confidence interval; CS, Calibration slope; EPICS, Emergency processes in clinical structures; EWS, Early warning score; HL GOF, Hosmer-Lemeshow goodness of fit; HOTEL, Hypotension, oxygen saturation, low temperature, ECG changes, loss of independence; MARS, Medical admissions risk system; MEWS, Modified early warning score; NEWS, National early warning score; NS, Not specified; OR, Odds ratio; PARIS, Systolic blood pressure, age, respiratory rate, loss of independence, peripheral oxygen saturation; PSS, Physiological scoring system; RAPS, Rapid acute physiology score; REMS, Rapid emergency medicine score; SCS, Simple clinical score; SE, Standard error; VSS, Vital sign score.*calculated from the available data.

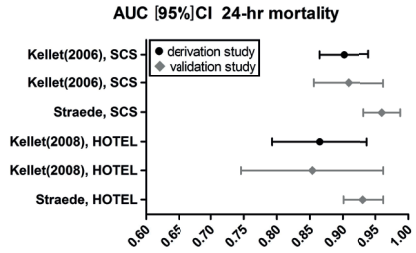


Figure 4.2.a: Discrimination performance of the models predicting 24-hour mortality

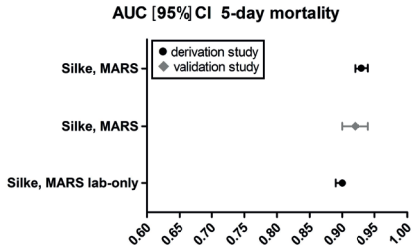


Figure 4.2.b: Discrimination performance of the models predicting 5-day mortality

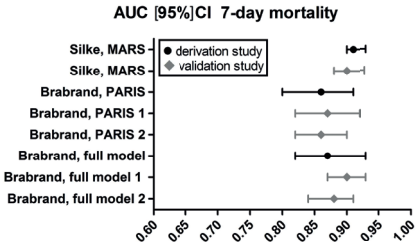


Figure 4.2.c: Discrimination performance of the models predicting 7-day mortality

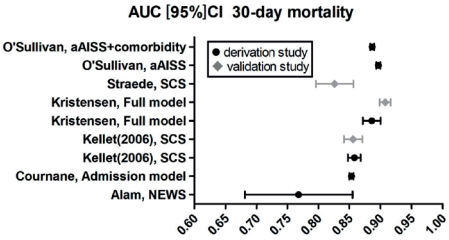


Figure 4.2.d: Discrimination performance of the models predicting 30-day mortality

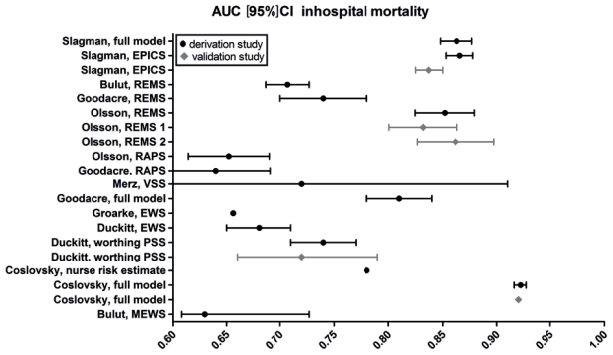


Figure 4.2.e: Discrimination performance of the models predicting inhospital mortality
In all figures a distinction is made between derivation studies and validation studies. Abbreviations: AUC, Area under the curve; CI, Confidence interval; SCS, Simple clinical score; HOTEL, Hypotension, oxygen saturation, low temperature, ECG changes, loss of independence; MARS, Medical admissions risk system; PARIS, Systolic blood pressure, age, respiratory rate, loss of independence, peripheral oxygen saturation; aAISS, adjusted Acute illness severity score; NEWS, National early warning score; EPICS, Emergency processes in clinical structures; REMS, Rapid emergency medicine score; RAPS, Rapid acute physiology score; VSS, Vital sign score; EWS, Early warning score; PSS, Physiological scoring system; MEWS, Modified early warning score.

DISCUSSION

In our systematic review we described models that predict short-term mortality of patients visiting the ED. To our knowledge, none of these models are currently implemented for mortality prediction in clinical practice. We assessed the methodological quality of the prediction models for discrimination, calibration and validation, where available. The discrimination of the models, presented by the AUC, ranged between average and excellent, with the majority having a good discriminatory performance. The MARS model had the highest performance, followed by the model by *Coslovsky et al.* and the SCS[132-134]. To determine the level of agreement between the expected and observed outcome, calibration is paramount. Calibration was assessed in seven studies, and was good in four, which shows that these models are suitable for validation[131-133, 135, 138, 141, 143].

Validation is needed before a model can be implemented in clinical practice, and external validation is preferred. Of the nine articles that described validation, only five used external validation. Internal validation was performed either using a split-sample, cross-validation or a bootstrap resampling technique. One study used bootstrap resampling[132], which is considered the best method, as it provides a true representation of the population without loss of patients[111]. In the studies that provided validation, the performance in all models was satisfying, and the highest performance was for the HOTEL score (AUC = 0.960)[136]. The PARIS model had insufficient calibration for the validation in one of the two validation cohorts[131]. This means that the model was not generalizable to one of the studied cohorts.

Table 4.5: Performance of the validated models

Study	Prediction model	Mortality, N(%)	Validation type	Discrimination, AUC [95% CI] / (±SE)	Calibration method	Calibration
24-hr mortality						
Kellett (2008)[135]	HOTEL	19 (0.6)	split-sample validation	0.854 [0.746 - 0.962]	NS	NS
Kellett (2006)[134]	SCS	19 (0.6)	split-sample validation	0.909 (± 0.027)	NS	NS
Straede[136]	SCS	26 (0.9)	external validation	0.960 [0.932 - 0.988]	HL GOF	χ^2 = 2.68 (0.998)
Straede[136]	HOTEL	26 (0.9)	external validation	0.931 [0.901 - 0.962]	HL GOF	χ^2 = 5.56 (0.234)
5-day mortality						
Silke[133]	MARS	171 (4.75)	external validation	0.92 [0.90 - 0.94]	HL GOF	χ^2 = 9.83 (0.278)
7-day mortality						
Brabrand[131]	Full model	1. 57 (2.0) 2. 111 (4.3)	external validation	1. 0.90 [0.87 - 0.93] 2. 0.88 [0.84 - 0.91]	HL GOF	P = 1. 0.75 2. 0.33
Brabrand[131]	PARIS	1. 57 (2.0) 2. 111 (4.3)	external validation	1. 0.87 [0.82 - 0.92] 2. 0.86 [0.82 - 0.90]	HL GOF	P = 1. 0.74 2. < 0.001
Silke[133]	MARS	216 (5.11)	external validation	0.90 [0.88 - 0.928]	HL GOF	χ^2 = 4.46 (0.814)
30-day mortality						
Kellett (2006)[134]	SCS	145 (4.5)	split-sample validation	0.856 (± 0.013)	NS	NS
Straede[136]	SCS	196 (6.4)	external validation	0.826 [0.774 - 0.879]	HL GOF	χ^2 =4.00 (0.947)
Kristensen[138]	Full model	234 (4.1)	cross-validation / external validation	0.908 [0.892 - 0.923]	BS	3.40 [3.08 - 3.72]
Inhospital mortality						
Coslovsky[132]	Full model	398 (4.6)	bootstrapping / split-sample validation	0.920	CS	CS 0.935 (split sample)
Coslovsky[132]	Nurse risk estimate model	398 (4.6)	bootstrapping	negligible difference	BS	negligible difference
Duckitt[141]	Worthing PSS	85 (8.0)	temporal validation	0.72 [0.66-0.79]	HL GOF	p = 0.565
Olsson[143]	REMS	285 (2.4)	split-sample validation	1. 0.832 (± 0.016) 2. 0.862 (± 0.018)	HL GOF	χ^2 = 1. 35.3 2. 31.7
Slagman[140]	EPICS	765 (2.1)	external validation	0.837 [0.825 - 0.850]	NS	NS

Performance of the different validation prediction tools for mortality, divided by the time-frame of mortality. Abbreviations: AUC, Area under the curve; BS, Brier score; CI, Confidence interval; CS, Calibration slope; EPICS, Emergency processes in clinical structures; HL GOF, Hosmer-Lemeshow goodness of fit; HOTEL, Hypotension, oxygen saturation, low temperature, ECG changes, loss of independence; MARS, Medical admissions risk system; NS, Not specified; PARIS, Systolic blood pressure, age, respiratory rate, loss of independence, peripheral oxygen saturation; PSS, Physiological scoring system; REMS, Rapid emergency medicine score; SCS, Simple clinical score; SE, Standard error; VSS, Vital sign score.

When we assessed the quality of the prediction models, the models by Coslovsky and Brabrand scored best with an overall low risk of bias on all the assessed domains[131, 132]. Only in the analysis domain high risk of bias was found, and this is explained by a lack of validation in these studies[86, 128, 137, 139, 142, 144, 145]. Some of the CHARMS criteria within the domains were missing in all studies. First, most studies lacked information on missing data or excluded patients with missing information. Excluding these patients, however, might limit not only the correctness, but also the usability of the model. In daily practice, the parameters of a model are not always available[146]. There are multiple options to address the issue of missing data. Missing values can be replaced by imputation, by the mean or by a normal value dependent on the type of missing data. It is also possible to assign a special category to missing values which correlates to a certain regression coefficient (thus mortality risk). Multiple imputation is considered to be the best method, since it gives reliable results without losing data[147, 148]. Unfortunately, just two articles used imputation methods to address missing values[131, 132].

Second, most articles did not describe loss of follow-up. However, it is questionable whether there is much influence of this loss of mortality data, since only short-term mortality (with a maximum of 30 days) was studied. Third, the number of variables that can be used in a model depends on the number of events (i.e. mortality) encountered in the study cohort. To limit overfitting in a model, there should be at least ten events per variable in order to include a parameter in a model[149]. The events per variable were only explicitly mentioned by *Brabrand et al.*[131]. Fourteen studies had enough events in relation to their number of variables[132-136, 139-145]. However, the studies of *Groarke* and *Alam* had less than ten events per variable[86, 128]. This could have been addressed by using a larger sample with more events.

The SCS, PARIS and full model of *Coslovsky* had a high performance of the model with good validation and low risk of bias[131, 132, 134]. However, a model should also be usable in clinical practice with a relevant predicted outcome. The relevance of a tool that predicts 24-hour mortality seems limited, as these patients presumably are more critically ill upon ED presentation.

A model that uses parameters with low inter-rater and intra-rater variability is reproducible and generally implementable. This can best be achieved by using objective measurements. Objective measurements also allow automatic calculation of the scores and the subsequent risks in an electronic patient file, and may even trigger alarms. Not all models met this prerequisite, as the SCS uses a patient's complaint of subjective breathlessness as a parameter[134].

For immediate and effective use a model should use parameters that are readily available and easily obtained. Eight of the models included in this systematic review used laboratory values as predictors, which entails a waiting time, and thereby delay in prediction[133, 137-140]. Six models used parameters requiring (collateral) history, such as loss of independence, confinement to bed, and comorbidities[131, 132, 134-136, 139]. In patients with an altered or lowered consciousness this information is not always available, which subsequently influences the results of the model. Furthermore, parameters such as the presence of seizures and APACHE II diagnostic category require both diagnostic testing, which takes time, and require patients to be in one of these diagnostic categories, which is not always the case. Therefore, these parameters are not applicable to all patients, and thus are the models not generalizable to the general ED population. We believe that in specific patient populations parameters like APACHE II diagnostic category will perform better than merely vital parameter. However, *Coslovsky et al.* showed that the effect of APACHE II category was less than vital signs, such as MAP (OR = 0.57 vs. OR = 0.93)[132]. Last, models with complex calculations require applications ("apps") or calculation programs, which could cause a delay in the risk calculation.

We found that for clinical use in the ED the RAPS, REMS, NEWS and EWS are most suited, since they use routinely acquired vital parameters, which meets the requirements of early, easily obtainable and objective predictors. However, the AUC of these models is lower compared to the PARIS and full score.

Future model development should ideally combine good model performance with clinical applicability. The use of a prospective cohort study design is warranted as it allows optimal predictor selection and outcome measurement[150]. Before implementation, a model should be externally validated to prove generalizability. Large datasets allow head-to-head comparison of multiple models in order to detect the best model. Most importantly, pre- and post-implementation measurements should be performed to determine if introduction of a mortality prediction tool leads to earlier identification of patients at risk, with subsequent faster initiation of treatment and a decrease in mortality as a final result. If introduction of a prediction tool at the ED finally does not yield these effects, its further implementation in clinical practice warrants of little use.

Our study has several strengths and limitations. Strengths of this study include the comprehensive search strategy and the methodological quality assessment with CHARMS, and both were executed by two researchers. There are also several limitations. First, in our review we identified highly heterogeneous studies, making it unfeasible to perform a formal meta-analysis. This heterogeneity makes it difficult to reliably rank different models, as the different models all have their merits and flaws. Second, selection bias

might be present. We attempted to minimize this risk by using two researchers to select the studies. Third, despite we only selected European studies, practice and organization between countries can differ. External validation might make these results more generalizable, however, as external validation was mostly done in the same country, these risks remain.

CONCLUSION

In conclusion, we provide an extensive overview of literature concerning prediction models for mortality in the ED for an unselected medical population. In general, the models performed well to excellent. Models with more and difficult obtainable parameters performed better. Most studies had bias due to the reporting of missing values, handling of missing data and lack of validation. These issues should be taken into account in future models. At this time, the PARIS model and the full-model of *Brabrand et al.* are the best performing models, however, these models require additional information such as loss of independence[131]. The EWS and NEWS use readily available parameters, but have lower performance. Therefore, the perfect model has yet to be developed.

Appendix 4.A: Search strategy for systematic review of the identification of prediction models for mortality in the emergency department

Embase.com
('prediction'/exp OR 'predictive value'/exp OR 'predictive validity'/exp OR 'prediction and forecasting'/de OR 'predictor variable'/exp OR (predict*):ab,ti) AND (mortality/exp OR survival/exp OR survivor/de OR 'fatality'/de OR (('intensive care'/exp OR 'intensive care unit'/exp) AND ('hospital admission'/exp OR 'hospitalization'/de)) OR (mortalit* OR surviv* OR fatal* OR ((admission* OR admit*) NEAR/3 (icu OR intensive-care*))):ab,ti) AND ('emergency care'/exp OR 'emergency patient'/exp OR 'emergency ward'/exp OR 'emergency health service'/exp OR ((emergen* NEAR/3 (ward* OR department* OR patient* OR service* OR admiss* OR admit* OR hospital* OR call*))):ab,ti) AND ('cohort analysis'/exp OR 'follow up'/exp OR 'longitudinal study'/de OR 'retrospective study'/de OR 'prospective study'/de OR 'evaluation study'/de OR model/de OR 'disease model'/de OR 'population model'/de OR 'process model'/de OR simulation/exp OR algorithm/de OR 'validation process'/exp OR 'sensitivity and specificity'/exp OR 'scoring system'/exp OR 'decision tree'/de OR (model OR simulat* OR cohort* OR (follow* NEXT/1 up*) OR followup* OR longitudinal* OR retrospectiv* OR prospectiv* OR evaluation* OR algorithm* OR validat* OR sensitivit* OR specificit* OR score* OR (decision NEXT/1 tree*)):ab,ti) NOT ([Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim) AND [english]/lim NOT ((child/exp OR childhood/exp OR adolescent/exp OR adolescence/exp) NOT (adult/exp OR adulthood/exp)) NOT (pediatrics/exp OR (picu OR nicu OR picus OR nicus OR pediatric* OR paediatric*):ab,ti)

Medline ovid
("Predictive Value of Tests"/ OR "Forecasting"/ OR (predict*).ab,ti.) AND (exp mortality/ OR mortality.xs. OR survival/ OR survivors/ OR (("Critical Care"/ OR "Intensive Care Units"/) AND ("Patient Admission"/ OR "hospitalization"/)) OR (mortalit* OR surviv* OR fatal* OR ((admission* OR admit*) ADJ3 (icu OR intensive-care*))).ab,ti.) AND ("Emergency Medical Services"/ OR "emergencies"/ OR exp "Emergency Service, Hospital"/ OR ((emergen* ADJ3 (ward* OR department* OR patient* OR service* OR admiss* OR admit* OR hospital* OR call*))).ab,ti.) AND (exp "cohort studies"/ OR "evaluation study"/ OR exp "Models, Statistical"/ OR "Computer Simulation"/ OR "Models, Theoretical"/ OR Algorithms/ OR "Validation Studies"/ OR exp "sensitivity and specificity"/ OR "Decision Trees"/ OR (model OR simulat* OR cohort* OR (follow* ADJ up*) OR followup* OR longitudinal* OR retrospectiv* OR prospectiv* OR evaluation* OR algorithm* OR validat* OR sensitivit* OR specificit* OR score* OR (decision ADJ tree*)).ab,ti.) NOT (letter OR news OR comment OR editorial OR congresses OR abstracts).pt. AND english.la. NOT ((exp child/ OR exp Infant/ OR adolescent/) NOT (exp adult/)) NOT (exp pediatrics/ OR (picu OR nicu OR picus OR nicus OR pediatric* OR paediatric*).ab,ti.)

Cochrane
((predict*):ab,ti) AND ((mortalit* OR surviv* OR fatal* OR ((admission* OR admit*) NEAR/3 (icu OR intensive-care*))):ab,ti) AND (((emergen* NEAR/3 (ward* OR department* OR patient* OR service* OR admiss* OR admit* OR hospital* OR call*))):ab,ti) AND ((model OR simulat* OR cohort* OR (follow* NEXT/1 up*) OR followup* OR longitudinal* OR retrospectiv* OR prospectiv* OR evaluation* OR algorithm* OR validat* OR sensitivit* OR specificit* OR score* OR (decision NEXT/1 tree*)):ab,ti)

Web-of-science
TS=(((predict*)) AND ((mortalit* OR surviv* OR fatal* OR ((admission* OR admit*) NEAR/3 (icu OR intensive-care*)))) AND (((emergen* NEAR/2 (ward* OR department* OR patient* OR service* OR admiss* OR admit* OR hospital* OR call*)))) AND ((model OR simulat* OR cohort* OR "Follow up" OR followup* OR longitudinal* OR retrospectiv* OR prospectiv* OR evaluation* OR algorithm* OR validat* OR sensitivit* OR specificit* OR score* OR (decision NEAR/1 tree*))) NOT ((child* OR infan* OR adolescen* OR newborn* OR neonat*) NOT (adult* OR older* OR elder* OR (aged NEAR/3 (person* OR patient*)))))

Google scholar
Prediction|predctive mortality|survival|fatal|fatality|"icu|care admission"|"admitted**icu|intensive"|"admitted*icu|intensive""emergency ward|department|patient|service" model|simulation|cohort|"follow up"|evaluation


Chapter
5

Predicting Mortality in Patients with Suspected Sepsis at the Emergency Department; A Retrospective Cohort Study Comparing qSOFA, SIRS and National Early Warning Score

Anniek Brink
Jelmer Alsma
Rob J.C.G. Verdonschot
Pleunie P.M. Rood
Robert Zietse
Hester F. Lingsma
Stephanie C.E. Schuit

PLoS One. 2019 Jan 25;14(1):e0211133.

ABSTRACT

Background

In hospitalised patients, the risk of sepsis-related mortality can be assessed using the quick Sepsis-related Organ Failure Assessment (qSOFA). Currently, different tools that predict deterioration such as the National Early Warning Score (NEWS) have been introduced in clinical practice in emergency departments (ED) worldwide. It remains ambiguous which screening tool for mortality at the ED is best. The objective of this study was to evaluate the predictive performance for mortality of two sepsis-based scores (i.e. qSOFA and Systemic Inflammatory Response Syndrome (SIRS)-criteria) compared to the more general NEWS score, in patients with suspected infection directly at presentation to the ED.

Methods

We performed a retrospective cohort study. Patients who presented to the ED between June 2012 and May 2016 with suspected sepsis in a large tertiary care center were included. Suspected sepsis was defined as initiation of intravenous antibiotics and/or collection of any culture in the ED. Outcome was defined as 10-day and 30-day mortality after ED presentation. Predictive performance was expressed as discrimination (AUC) and calibration using Hosmer-Lemeshow goodness-of-fit test. Subsequently, sensitivity, and specificity were calculated.

Results

In total 8,204 patients were included of whom 286 (3.5%) died within ten days and 490 (6.0%) within 30 days after presentation. NEWS had the best performance, followed by qSOFA and SIRS (10-day AUC: 0.837, 0.744, 0.646, 30-day AUC: 0.779, 0.697, 0.631). qSOFA (≥ 2) lacked a high sensitivity versus SIRS (≥ 2) and NEWS (≥ 7) (28.5%, 77.2%, 68.0%), whilst entailing highest specificity versus NEWS and SIRS (93.7%, 66.5%, 37.6%).

Conclusions

NEWS is more accurate in predicting 10- and 30-day mortality than qSOFA and SIRS in patients presenting to the ED with suspected sepsis.

INTRODUCTION

Sepsis is a syndrome characterised by both signs of infection and manifestations of a systemic host response[55]. Sepsis is the primary cause of mortality from infection. The definition of sepsis has changed throughout the last decades. In February 2016 the Third International Consensus Definition for Sepsis (Sepsis-3) replaced the Sepsis-2 definition dating from 2001[52, 53, 55]. Sepsis is currently defined as a "life-threatening organ dysfunction caused by a dysregulated host response to infection", in which organ dysfunction is represented by an increase of at least two points in the Sequential Organ Failure Assessment (SOFA) score[55]. The Systemic Inflammatory Response Syndrome (SIRS) score, which was part of the definition in Sepsis-1 and -2, has been abandoned.

The quick Sepsis-related Organ Failure Assessment (qSOFA) was introduced with the new Sepsis-3 definition[151]. However, not all medical societies support this new definition[152, 153]. The qSOFA consists of three parameters (i.e. low systolic blood pressure (≤ 100 mmHg), tachypnoea (≥ 22/minute) and altered mental status (Glasgow Coma Scale (GCS) < 15 / AVPU < Alert)), with a maximum score of three points. qSOFA is a bedside prompt to identify patients with a suspected infection who are at greater risk for a poor outcome. It is a simplified score based on the SOFA score. Early identification of these patients potentially results in earlier adequate treatment and a decrease in mortality. qSOFA aims to prognosticate the course of sepsis and intends to predict sepsis-related mortality and adverse events; a score of two points or higher gives a three to 14-fold increase in inhospital mortality[151]. The qSOFA score is claimed to be more accurate than SOFA in departments outside the intensive care unit (ICU), however the use of qSOFA in the emergency department (ED) is questionable[151, 154-157]. The authors of Sepsis-3 also consider qSOFA as a prompt to identify possible infection[55].

In many patients admitted to the ED with sepsis the severity of their illness is not directly clear. The presence of a life-threatening infection can easily be overlooked. The use of screening tools in the ED can aid in early recognition of patients with sepsis, resulting in early initiation of effective and complete treatment. This requires screening tools with a high sensitivity. SIRS has been criticized for being too sensitive, while lacking specificity in recognizing sepsis, and it is therefore not an ideal screening tool. As qSOFA performed better than SIRS in hospitalised patients, it has been proposed that qSOFA is preferred to SIRS. Alternatively, early warning scores, such as the National Early Warning Score (NEWS), are already recommended for use in the ED, and should therefore also be considered[158]. NEWS was introduced in 2012 by the Royal College of Physicians, who aimed to provide a standardised early warning score. This score is used for early detection of patients at risk for deterioration but is not specific for sepsis. NEWS comprises of

seven parameters (i.e. respiratory rate, oxygen saturation, supplemental oxygen, body temperature, systolic blood pressure, heart rate, AVPU score) with a maximum of twenty points. In clinical practice cut-off values of 1-4, 5-6 and ≥ 7, respectively for low, medium and high risk are used. NEWS was primarily developed for use on the wards, however NEWS was also tested for use in the ED and in the prehospital setting[159, 160]. For use in the ED a cut-off value of ≥ 7 is suggested.

The aim of this study was to determine the prognostic value of qSOFA in predicting mortality in comparison to SIRS and NEWS in patients presenting to the ED with suspected sepsis.

METHODS

Study design and setting

This was a retrospective cohort study nested in a large anonymous database of patients visiting the ED of the Erasmus University Medical Center, Rotterdam, the Netherlands (Erasmus MC), which is the largest tertiary referral centre in The Netherlands. The ED is an open access department with approximately 30,000 annual visits. Patients are strongly encouraged to see a general practitioner before visiting the ED. The database of the ED consists of all patients presenting to the ED. This database holds information of patients from January 2012 and onwards, on both clinical and vital parameters, laboratory results, other diagnostic procedures and treatments. The data was extracted from the electronic health records every two weeks through May 2017. Random samples were manually checked for concordance.

Selection of participants

In our consecutive cohort, we included patients with suspected sepsis visiting the ED between June 1st 2012 and May 31st 2016. Suspected sepsis was defined as either the initiation of non-prophylactic intravenous antibiotic therapy during their ED visit or the collection of any culture (i.e. blood cultures, urine cultures, wound cultures, throat swabs, sputum cultures and cultures of cerebrospinal fluid) or viral diagnostics (i.e. polymerase chain reaction (PCR) on blood and stool samples, on throat swabs and on cerebrospinal fluids) during the index visit. Rapid diagnostic testing for viral or bacterial infections was not possible during the study period. Patients who presented with symptoms directly related to trauma were excluded. A comprehensive search in the database identified all patients who met this definition.

Measurements and Outcomes

Demographic data (i.e. age, sex), vital parameters (i.e. blood pressure, body temperature, respiratory rate, peripheral oxygen saturation, consciousness level according to AVPU scale or GCS), laboratory testing performed, acuity level according to Manchester Triage System (MTS) category, and supplemental oxygen therapy were derived from the database.

The AVPU scale is a system to score the mental status and is an acronym of 'Alert, Verbal, Pain, Unresponsive'[161]. When AVPU was not scored, GCS was used, and vice versa. Only the first vital parameters were retrieved as the aim of the study was to assess the ability of the different prompts to screen for short-term mortality at ED presentation. White blood cell count was retrieved for all patients when available. Data on all-cause mortality was obtained from patient records and 10- and 30-day mortality was calculated. Mortality data was retrieved from the patient records, which are linked to municipal mortality data. Subsequently, we assessed whether mortality was directly sepsis-related or not.

We calculated qSOFA, SIRS and NEWS and formed groups using cut-off values most indicative for poor outcome (qSOFA ≥ 2, SIRS ≥ 2, and NEWS ≥ 7)(*Table 5.1*)[52, 151, 158]. The Medical Ethics Committee of the Erasmus MC reviewed the study and deemed exempt.

Statistical analysis

Data was summarized using mean, median, interquartile range (IQR) and standard deviation (SD) when appropriate. Missing or clinically implausible data was replaced by multiple imputation. This method is valid even when large sets of data are missing[162]. Missing values within the parameters were imputed five times using non-missing parameters. Furthermore, imputation was based on a distribution of the observed data to preclude that implausible values would replace the missing value. After imputation, five complete datasets were available. In each dataset the SIRS, qSOFA and NEWS scores were recalculated using the imputed variables. Whenever possible, results were pooled. When pooling was not possible, single imputation was used. The primary outcome was all-cause mortality within 10- and 30-days after ED presentation.

Patient characteristics were compared using the two-sampled t-test, Mann-Whitney U test, and chi-squared test based on the distribution of the data. Univariate regression analysis was used for association between the different parameters and 10- and 30-day mortality to determine which variable is the best predictor. This predictor is characterized by the largest LRχ^2 and a high explained variance (i.e. R^2 close to one).

Table 5.1: Variables within NEWS, qSOFA and SIRS criteria

	NEWS[a]							qSOFA[b]			SIRS[c]		
	3	2	1	0	1	2	3	1	0	1	1	0	1
Body temperature (°C)	≤ 35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥ 39.1					< 36.0	36.0 - 38.0	> 38.0
Heart rate (bpm)	≤ 40		41 - 50	51 - 90	91 - 110	111 - 130	≥ 131					≤ 90	> 90
Systolic blood pressure (mmHg)	≤ 90	91 - 100	101 - 110	111 - 219			≥ 220	≤ 100	> 100				
Respiratory rate (per minute)	≤ 8		9 - 11	12 - 20		21 - 24	≥ 25		< 22	≥ 22		≤ 20	> 20
Oxygen saturation (%)	≤ 91	92 - 93	94 - 95	≥ 96									
Supplemental oxygen		Yes		No									
AVPU score / GCS				A / 15			V,P,U / < 15		A / 15	V,P,U / < 15			
WBC (*10^9/L)											≤ 4.0	4.0 - 12.0	> 12.0

Variables within the National Early Warning Score, quick Sepsis-related Organ Failure Assessment and Systemic Inflammatory Response Syndrome criteria. Each variable is measured and summed up. [a]NEWS ranges from 0 to 20, wherein 1 to 3 points are given for aberrant values in the following variables: body temperature, heart rate, systolic blood pressure, respiratory rate, oxygen saturation, supplemental oxygen and AVPU score. [b]qSOFA ranges from 0 to 3, in which 1 point is assigned to abnormal values in the following variables: systolic blood pressure, respiratory rate and AVPU score. [c]SIRS ranges from 0 to 4 points, wherein 1 point is allocated to aberrant values in the following variables: body temperature, heart rate, respiratory rate and WBC. The total score within NEWS, qSOFA, and SIRS corresponds with a risk for mortality. Abbreviations: NEWS, national early warning score; qSOFA, quick sepsis-related organ failure assessment; SIRS, systemic inflammatory response syndrome; °C, degrees Celsius; bpm, beats per minute; mmHg; millimetre of mercury; AVPU, alert, verbal, pain, unresponsive; WBC, white blood cell count.

Logistic regression was used to obtain the odds for 10- and 30-day mortality based on individual scores. The predictive performances of qSOFA, SIRS, and NEWS were expressed as discrimination (area under the Receiver Operating Characteristic-curve) and calibration. Calibration represents how mortality predictions resemble the observed mortality, which was measured by the Hosmer-Lemeshow goodness-of-fit test and expressed as a χ^2-value and accessory p-value. Subsequently, sensitivity, specificity and positive- and negative predictive values were calculated for the different cut-off points. The Youden's J statistic was calculated to assess the optimal cut-off point for the different scores. A p-value < 0.05 was considered statistically significant. Analyses were undertaken using Statistical Package for the Social Science (SPSS) version 21 and R statistics version 3.1.3. (2015-03-09).

RESULTS

Patient characteristics

A total of 120,177 ED visits in 75,428 unique patients were recorded between June 1st 2012 and May 31st 2016. 21,326 patient records were excluded as their ED visits were related to trauma, leaving 54,102 patients for analysis. 3,351 patients received intravenous antibiotic therapy in the ED. Bacterial cultures and viral diagnostics were collected from 7,302 patients during their ED visit. In total, 8,204 patients were analysed (*Figure 5.1*). The majority of patients were male (55.9%), and the median age was 57.0 (IQR 41.0 - 67.0). In total, 74.6% of patients were hospitalised (*Table 5.2*). 10-day and 30-day mortality was 3.5% (286) and 6.0% (490), respectively. Of the 490 deceased patients, 64,7% died in the hospital. Patients who died were significantly older, and had higher heart rates, lower systolic blood pressures, lower oxygen saturation and higher respiratory rates during ED presentation. 18,4% of the deceased patients had positive cultures. The cause of death could be retrieved from the patient records in all 490 deceased patients. In 63.4% of patients their death was directly related to sepsis.

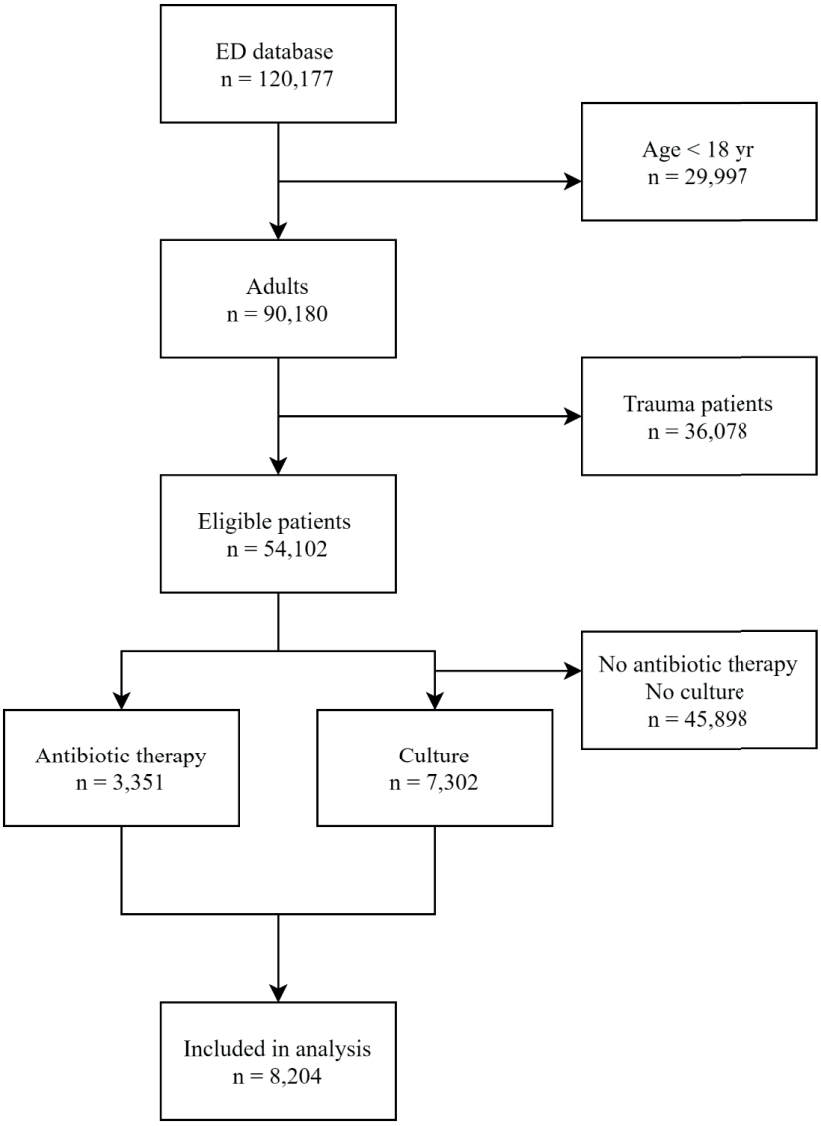


Figure 5.1: Subject inclusion flowchart. Flowchart of patients who met inclusion/exclusion criteria.

Performance of the models

Univariate regression analysis showed that oxygen therapy during ED presentation - a variable within NEWS - was the best predictor for mortality (LRχ^2 = 335.73), although the explained variation was low (r^2 = 0.110). Other strong predictors included systolic blood pressure and mental status (*Table 5.3*).

Table 5.2: Patient characteristics

	N (% missing)	All patients	Died within ten days	Died within 30 days	Alive	p-value
N (%)		**8,204**	**286 (3.5)**	**490 (6.0)**	**7,714 (94.0)**	
Male, N (%)	8,204 (0)	4,581(55.8)	182 (63.6)	321 (65.5)	4,260 (55.2)	< 0.0001*
Age, median (IQR)	8,204 (0)	57.0 (41 - 68)	68.0 (58.75 - 78)	67.0 (58 - 77.25)	56.0 (41-67)	< 0.0001†
Body temperature in °C, mean (SD)	7,945 (3.2)	37.6 (1.3)	36.9 (1.7)	37.2 (1.5)	37.7 (1.2)	< 0.0001‡
HR in bpm, mean (SD)	7,858 (4.2)	97.9 (21.4)	103.7 (26.5)	104.9 (26.1)	97.5 (21.0)	< 0.0001‡
SBP in mmHg, mean (SD)	7,764 (5.4)	131.7 (26.1)	119.6 (36.2)	121.3 (34.0)	132.3 (25.4)	< 0.0001‡
RR per minute, mean (SD)	4,796 (41.5)	21.3 (8.5)	25.0 (9.1)	24.5 (9.1)	21.0 (8.3)	< 0.0001‡
Oxygen saturation in %, mean (SD)	7,578 (7.6)	96.0 (3.6)	93.9 (5.9)	93.9 (5.6)	96.2 (3.4)	< 0.0001‡
AVPU, N (%)	6,643 (19.0)					< 0.0001§
Alert		6,104 (91.9)	152 (64.7)	291 (72.6)	5,813 (93.1)	
Verbal		385 (5.8)	39 (16.6)	57 (14.2)	328 (5.3)	
Pain		69 (1.0)	12 (5.1)	16 (4.0)	53 (0.8)	
Unresponsive		85 (1.3)	32 (13.6)	37 (9.2)	48 (0.8)	
Supplemental oxygen, N (%)	8,204 (0)	2,472 (30.1)	223 (78.0)	338 (69.0)	2,134 (27.7)	< 0.0001*
Laboratory testing performed, N (%)	8,204 (0)	6,980 (86.9)	251 (87.8)	437 (89.2)	6,690 (86.7)	0.118*
WBC in $*10^9$/L, mean (SD)	7,036 (14.2)	11.84 (12.88)	17.03 (30.70)	15.37 (24.22)	11.58 (11.58)	< 0.0001‡
SIRS ≥ 2, N (%)	4,387 (46.5)	2,940 (67.0)	178 (62.2)	298 (78.6)	2,642 (65.9)	< 0.0001*
qSOFA ≥ 2, N (%)	4,318 (47.4)	369 (4.5)	59 (20.6)	87 (17.8)	282 (7.0)	< 0.0001*
NEWS ≥ 7, N (%)	4,243 (48.3)	1,895 (44.7)	135 (77.1)	212 (70.0)	1,683 (42.7)	< 0.0001*
MTS, N (%)	7,786 (5.1)					< 0.0001§
Immediate		168 (2.2)	47 (18.2)	53 (11.8)	115 (1.6)	
Very urgent		1,002 (12.9)	87 (33.7)	148 (32.9)	854 (11.6)	
Urgent		5,451 (70.0)	115 (44.6)	230 (51.1)	5,221 (71.2)	
Standard		1,144 (14.7)	9 (3.5)	19 (4.2)	1,125 (15.3)	
Non urgent		16 (0.2)	0 (0.0)	0 (0.0)	16 (0.2)	
Admission, N (%)	8,204 (0)	6,117 (74.6)	273 (95.5)	455 (92.9)	5,662 (73.4)	< 0.0001*

Patient characteristics. Abbreviations: N, number; SBP, systolic blood pressure; RR, respiratory rate; HR, heart rate; AVPU, Alert, Verbal, Pain, Unresponsive), WBC white blood cell count; SIRS, systemic inflammatory response syndrome; qSOFA, quick sepsis-related organ failure assessment; NEWS, national early warning score; MTS, Manchester Triage System; IQR, interquartile range (25-75 percentile); SD, standard deviation; bpm, beats per minute; mmHg, millimeter of mercury; L, liter; °C, degrees Celsius. *Chi-squared test, †median test, ‡independent samples t-test, §Mann-Whitney U test.

Table 5.3: Univariate regression on the outcome 30-day mortality

		$LR\chi^2$	R^2
SIRS	Body temperature	0.51	0.000
	Heart rate	24.05	0.008
	Respiratory rate	28.13	0.013
	WBC	60.50	0.022
qSOFA	Respiratory rate	22.50	0.010
	Systolic blood pressure	133.49	0.045
	AVPU	142.03	0.060
NEWS	Oxygen therapy	335.73	0.110
	Oxygen saturation	44.54	0.016
	Respiratory rate	30.32	0.014
	Body temperature	17.13	0.006
	Systolic blood pressure	103.87	0.035
	Heart rate	43.04	0.015
	AVPU	144.17	0.059

30-day mortality univariate regression. The best parameter in the univariate model has the highest likelihood function ($LR\chi^2$). R^2 is the proportion of the variance in outcome 30-day mortality explained by the univariate model.

NEWS performed substantially better than qSOFA and SIRS in predicting both 10-day mortality (AUC [95% CI]: 0.837 [0.812 - 0.861], 0.744 [0.708 - 0.78] and 0.646 [0.613 - 0.679] respectively) and 30-day mortality (0.779 [0.755 - 0.804], 0.697 [0.667 - 0.726] and 0.631 [0.605 - 0.656] respectively) (*Figure 5.2* and *5.3*).

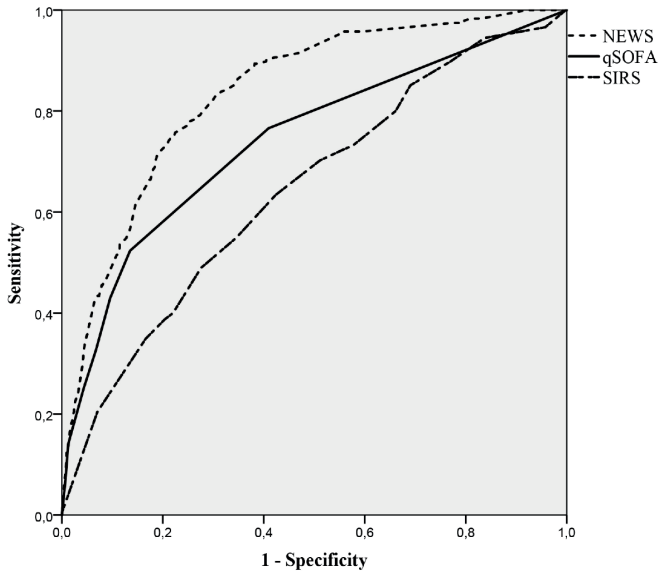


Figure 5.2: ROC Curve 10-day mortality.

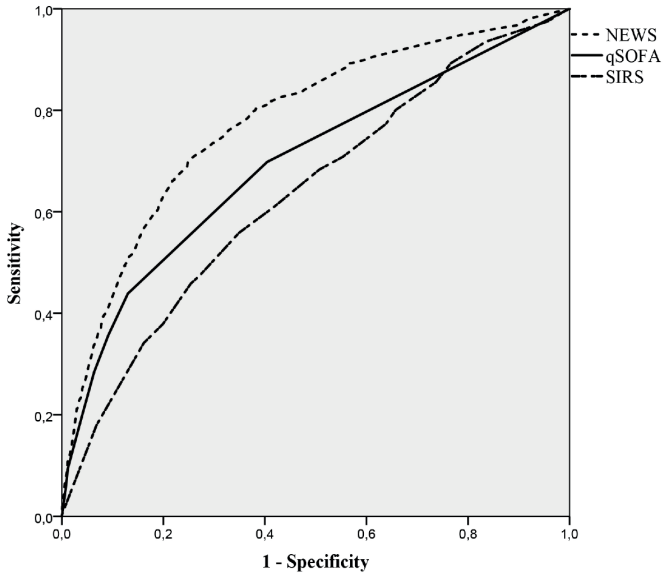


Figure 5.3: ROC Curve 30-day mortality.

Calibration for NEWS showed a $\chi^2 = 10.743$ and p-value = 0.217, compared to $\chi^2 = 6.915$ and p-value = 0.032 for qSOFA, and $\chi^2 = 22.827$ and p-value = 0.004 for SIRS. The non-significant p-value indicates that the mortality rates between the observed and the predicted values were statistically equivalent.

qSOFA showed the highest specificity, followed by NEWS and SIRS. Sensitivity was highest in SIRS, followed by NEWS and qSOFA. Using Youden's J statistic, the optimal cut-off points for 10-day mortality were qSOFA ≥ 1, SIRS ≥ 2 and NEWS ≥ 7, and for 30-day qSOFA ≥ 1, SIRS ≥ 3 and NEWS ≥ 7 (*Table 5.4*).

Table 5.4: Sensitivity, specificity, PPV, NPV and Youden's index for different cut-off values for 10- and 30-day mortality

10-day mortality	Sensitivity [95% CI] (%)	Specificity [95% CI] (%)	PPV (%)	NPV (%)	Youden's index
SIRS					
≥ 1	98.0 [95.5 - 99.2]	12.2 [11.5 - 12.9]	3.9	99.4	0.102
≥ 2‖	80.4 [75.3 - 84.9]	37.3 [36.2 - 38.4]	4.4	98.1	0.177¶
≥ 3	50.4 [44.4 - 56.3]	67.0 [66.0 - 68.0]	5.2	97.3	0.174
4	15.0 [11.1 - 19.7]	90.8 [90.2 - 91.4]	5.5	96.7	0.058
qSOFA					
≥ 1	77.2 [72.0 - 82.0]	59.1 [58.0 - 60.2]	6.5	98.6	0.362¶
≥ 2‖	33.1 [27.8 - 39.0]	93.3 [92.7 - 93.8]	15.3	97.4	0.264
3	7.8 [4.9 - 11.4]	99.3 [99.1 - 99.5]	28.2	96.7	0.071

Table 5.4: Sensitivity, specificity, PPV, NPV and Youden's index for different cut-off values for 10- and 30-day mortality (*continued*)

10-day mortality	Sensitivity [95% CI] (%)	Specificity [95% CI] (%)	PPV (%)	NPV (%)	Youden's index
NEWS					
≥ 3	98.3 [96.0 - 99.4]	17.8 [17.0 - 18.7]	4.2	99.7	0.161
≥ 4	94.5 [91.1 - 96.8]	26.0 [25.0 - 27.0]	4.5	99.2	0.205
≥ 5	89.1 [85.0 - 92.5]	42.1 [41.0 - 43.2]	5.3	99.1	0.312
≥ 6	82.1 [77.2 - 86.4]	57.0 [56.0 - 58.1]	6.5	98.9	0.391
≥ 7‖	76.3 [70.9 - 81.0]	65.9 [64.8 - 66.9]	7.6	98.7	0.421¶
≥ 8	59.6 [53.5 - 65.2]	77.1 [76.2 - 78.0]	8.7	98.1	0.367
≥ 9	45.8 [40.0 - 51.8]	84.0 [83.2 - 84.8]	9.5	97.7	0.298
≥ 10	35.1 [29.4 - 40.8]	89.4 [88.7 - 90.1]	10.8	97.4	0.245
≥ 11	22.8 [18.0 - 28.0]	94.5 [94.0 - 95.0]	13.2	97.1	0.173
≥ 12	9.4 [6.3 - 13.4]	98.3 [98.0 - 98.6]	17.3	96.7	0.078
≥ 13	9.4 [6.3 - 13.4]	98.3 [98.0 - 98.3]	17.3	96.7	0.078
≥ 14	4.2 [2.2 - 7.2]	99.3 [99.1 - 99.5]	17.9	96.6	0.035
≥ 15	1.2 [0.2 - 3.0]	99.7 [99.6 - 99.8]	14.1	96.5	0.009
≥ 16	0.3 [0.0 - 1.9]	99.9 [99.8 - 100.0]	15.4	96.5	0.003
30-day mortality	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**	**Youden's index**
SIRS					
≥ 1	96.3 [94.3 - 97.8]	12.4 [11.7 - 13.2]	6.5	98.1	0.087
≥ 2‖	77.2 [73.2 - 80.8]	37.6 [36.5 - 38.7]	7.3	96.3	0.148
≥ 3	48.1 [43.7 - 52.7]	67.3 [66.2 - 68.4]	8.5	95.3	0.154¶
4	14.9 [11.9 - 18.4]	90.9 [90.2 - 91.5]	9.4	94.4	0.058
qSOFA					
≥ 1	69.9 [65.7 - 74.0]	59.5 [58.0 - 60.2]	10.0	96.9	0.294¶
≥ 2‖	28.5	93.7 [92.7 - 93.8]	22.6	95.3	0.222
3	5.5 [3.7 - 7.9]	99.3 [99.1 - 91.5]	34.0	94.2	0.048
NEWS					
≥ 3	95.6 [93.3 - 97.1]	18.1 [17.2 - 19.0]	7.0	98.5	0.137
≥ 4	90.6 [87.7 - 93.0]	26.3 [25.3 - 27.3]	7.3	97.8	0.169
≥ 5	83.0 [79.4 - 86.3]	42.5 [41.4 - 43.6]	8.5	97.5	0.255
≥ 6	75.5 [71.4 - 79.3]	57.6 [56.5 - 58.7]	10.2	97.3	0.33
≥ 7‖	68.0 [63.6 - 72.1]	66.5 [65.4 - 67.6]	11.5	97.0	0.345¶
≥ 8	55.0 [50.6 - 59.6]	77.8 [76.8 - 78.7]	13.7	96.4	0.328
≥ 9	42.0 [37.6 - 46.5]	84.5 [83.7 - 85.3]	14.9	95.8	0.266
≥ 10	31.3 [27.1 - 35.5]	89.8 [89.1 - 90.5]	16.5	95.3	0.211
≥ 11	20.9 [17.3 - 24.7]	94.8 [94.3 - 95.3]	20.7	94.9	0.158
≥ 12	14.7 [11.7 - 18.1]	96.8 [96.4 - 97.2]	22.6	94.6	0.114
≥ 13	8.1 [5.9 - 11.0]	98.5 [98.2 - 98.8]	25.3	94.3	0.066

Table 5.4: Sensitivity, specificity, PPV, NPV and Youden's index for different cut-off values for 10- and 30-day mortality (*continued*)

10-day mortality	Sensitivity [95% CI] (%)	Specificity [95% CI] (%)	PPV (%)	NPV (%)	Youden's index
≥ 14	3.9 [2.4 - 6.0]	99.4 [99.2 - 99.6]	28.5	94.1	0.033
≥ 15	1.0 [0.3 - 2.4]	99.7 [99.6 - 99.8]	20	94.0	0.007
≥ 16	0.4 [0.1 - 1.5]	99.9 [99.8 - 100.0]	11.25	94.1	0.004

Sensitivity, specificity, positive predictive value, negative predictive value and Youden's index for different cut-off values for 10- and 30-day mortality, respectively. ‖ are the predefined cut-off values which are most indicative for a poor outcome. ¶ representing the optimal cut-off points. Abbreviations: PPV, positive predictive value; NPV, negative predictive value; SIRS, systemic inflammatory response syndrome; qSOFA, quick sepsis-related organ failure assessment; NEWS, national early warning score.

DISCUSSION

In this retrospective observational study of patients visiting the ED with a suspected sepsis we found that NEWS was superior to qSOFA and SIRS in predicting 10- and 30-day mortality for both discrimination and calibration. The different prompts all have different sensitivities and specificities for mortality. qSOFA has the highest specificity and lowest sensitivity, SIRS has the lowest specificity and highest sensitivity. NEWS has both an intermediate sensitivity and specificity, but is the best overall predictor in distinguishing high risk from low risk patients. NEWS has a lower sensitivity resulting in a significant number of false negatives, i.e. not all the patients who eventually died were identified with NEWS. NEWS was the only model with a good agreement between the expected and observed outcomes, i.e. calibration. However, none of the prediction models succeeded to fulfil all performance assessments, which would ideally be the case. Subsequent measurements of NEWS (e.g. hourly) will potentially identify patients who deteriorate during the stay in the ED and may improve sensitivity. We conclude that at presentation to the ED NEWS can be used as an alternative screening tool for patients with suspected sepsis who are at risk for deterioration, multi-organ failure, and subsequently death.

Our findings support the increasing data that suggests that the NEWS score is a useful screening tool in the ED, although its use has not fully been validated in the ED setting. *Jo et al.* studied the NEWS combined with serum lactate in predicting mortality in the general adult ED population and found an excellent discrimination (AUC = 0.96) for predicting two-day mortality[163]. The NEWS score as measured in the prehospital setting showed good correlation ($p < 0.001$) with hospital disposition[164]. Our study confirms the findings by *Churpek et al.* which support the introduction of the NEWS score in the ED. However, they studied patients outside the ICU and not only ED patients. And they primarily measured the performance of the different prompts based on the worst

vital signs. NEWS had the highest performance in predicting inhospital mortality in ED patients compared to qSOFA and SIRS (AUC = 0.77, AUC = 0.69 and AUC = 0.65 respectively). We used vital parameters at presentation in the ED and found similar results. In the *Churpek et al.* study a NEWS threshold of ≥ 7 is suggested. This threshold is also recommended by the Royal College of Physicians[158]. We were able to confirm this threshold using our data. In a cohort study by *Sbiti-Rohr et al.* in patients with community-acquired pneumonia, the NEWS score in the ED was significantly higher for those who died within 30 days after presentation than for survivors[165]. These results are similar to a study of patients presenting to the ED with acute dyspnoea; survivors had significantly lower NEWS scores at ED presentation[166].

The NEWS was also studied in patients suspected of sepsis. *Corfield et al.* found that an increased NEWS on arrival at the ED was associated with mortality in patients who met the sepsis criteria as defined by *Bone et al.* (odds ratio 1.95 to 5.64)[167].

Most prediction scores include measurements which are subject to interpretation. A study on the interrater agreement of GCS assessed at the ED yielded low agreement[168]. *Semler et al.* showed that in hospitalised patients recorded respiratory rates were higher than directly observed measurements. Also, the recorded rates were more likely to be 18 or 20 breaths/minute[169]. We expect that parameters that are not acquired automatically are subject to confounding by disease severity and were more likely to be measured and noted when one would expect a deviant result[170, 171]. Therefore, for the proper use of the NEWS, qSOFA and SIRS these measurements should be routinely performed in a structural way.

Specific scoring systems are used as an alternative to the NEWS to predict sepsis-related mortality in ED patients. The SIRS criteria, as introduced by Bone in 1992, were studied as a prediction tool for mortality and most studies show that an increase in SIRS items reflects an increased risk of mortality, ranging from 1.4% to 12% when no SIRS criteria were met and increasing to approximately 36% for four SIRS items[172, 173]. In Sepsis-3, the qSOFA was introduced as a simple tool to detect deterioration and predict mortality in departments outside the ICU. Simultaneously, SIRS criteria were abandoned from the new sepsis definition after criticism of its low specificity. The qSOFA ≥ 2 resembles a three to 14-fold increase in mortality risk[151].

qSOFA has been challenged as a prompt in the ED to identify patients with an increased risk for sepsis-related mortality ever since its introduction. Despite a high specificity (84 - 96%), the qSOFA has low sensitivity (13 - 53%)[154, 174]. This low sensitivity can be explained by the fact that the qSOFA is composed of vital parameters representing late

symptoms of deterioration (e.g. altered mental status due to inadequate perfusion of the brain)[153, 175]. In addition, qSOFA was derived in a cohort of critically ill patients, in which 11% of the patients were admitted in the ICU[151]. These patients represent a selected population compared to all patients who visit the ED, therefore, selection bias may be present. Furthermore, qSOFA was developed on the most aberrant results in serial vital parameter measurements. This approach may ameliorate the ability to predict mortality, but it restricts the utility as a prompt for early identification of patients at risk directly at ED presentation. All these arguments mainly affect the sensitivity and can influence the predictive performance of qSOFA. To increase sensitivity, *Park et al.* proposed the use of the qSOFA cut-off point of ≥ 1 instead of 2 for patients in the ED, resulting in an increase in sensitivity from 53.0% to 82.0%. This is in line with our findings. Changing the cut-off to 1 would increase the usability of qSOFA as a screening tool at cost of specificity. However, NEWS still has a higher sensitivity and a better predictive performance.

Strengths and limitations

This study has a number of strengths and limitations. The major strength of our study is that we used a large consecutive dataset with many relevant parameters directly derived from electronic patient records with mortality data directly acquired from municipality data.

Our study also has several limitations. The first limitation of this study is its retrospective design using data from a single tertiary care centre. In our centre we treat many patients with congenital and acquired immunodeficiencies (e.g. patients with organ or bone marrow transplantation, chemotherapy), which may limit the generalizability. The database contained missing values, which were replaced by multiple imputation. Multiple imputation has also been used in other sepsis-related studies[151, 154, 176]. Respiratory rate was most frequently missing and, as mentioned earlier, availability of respiratory rate might be an indicator of confounding by indication, as it is more often measured in patients who are deemed more critically ill[171]. A second limitation is the definition of the study population. As there is no gold standard for defining an infection, the study population was difficult to determine. We based our inclusion criteria on the definition of *Seymour et al.*[151], but modified the criteria to incorporate the largest group of patients who were suspected for infection and at risk for sepsis. Both microbial diagnostics and initiation of antibiotics were used as a proxy for a clinically suspected sepsis. These inclusion criteria could possibly bias against people with viral disease, as no antibiotics given and cultures are not routinely performed. However, in the most critically ill patients cultures are taken and antibiotics are started empirically in clinical practice, regardless of the suspected pathogen (e.g. virus, bacteria). Furthermore, we also included viral cultures

such as throat swabs and stool cultures, but these were a minority as compared to blood cultures (289 and 46 vs. 6552). Therefore, the chance of bias due viral sepsis is limited.

Last, to determine the best screening tool at presentation in the ED, we chose to use only the first recorded vital signs for calculation of NEWS, qSOFA and SIRS. We are aware that rapid changes in vital parameters could be indicative for a higher risk for mortality and that people may deteriorate during their ED visit. However, the duration of ED stay is intended to be very limited. Choosing to only use the first vital parameters may limit the predictive ability of the different models. However, in clinical practice the first vital parameters are used to determine the severity of the patient's condition and, therefore, to triage patients in urgent and non-urgent. Using first available parameters in this study actually reflects clinical practice and in our opinion is a valid method to test predictive performance upon ED presentation, with results comparable to using the worst vital parameters[177].

CONCLUSION

In conclusion, the NEWS is more accurate in predicting 10- and 30-day mortality than qSOFA and SIRS in patients suspected of sepsis on initial presentation to the ED. Our finding suggests that the introduction of the NEWS in the ED with subsequent measurements should be further studied. This will potentially aid the early detection of all patients at risk for deterioration in the ED including those at risk of sepsis-related mortality.

chapter 6

Performance of Early Warning Scores Predicting 30-day Mortality in Older Emergency Department Patients

Anniek Brink
Laura C. Blomaard
Jelmer Alsma
Romy Schuttevaer
Jacinta A. Lucke
Bas de Groot
Sander Anten
Robert Zietse
Simon Mooijaart
Hester F. Lingsma
Stephanie C.E. Schuit.

Submitted.

ABSTRACT

Background

Early warning scores (EWS) are designed to assess acuity in the emergency department (ED), but are known to perform poor in older patients presenting to the ED. Most likely this is the result of unspecific complaints, presence of comorbidities and frailty. We aim to study the predictive performance of EWS in the older ED population, and whether they can be improved by including additional predictors.

Methods

We conducted a secondary analysis of a multicentre prospective cohort study in older (≥ 70 years) ED patients. Four different EWS were calculated at ED triage. We refitted the coefficients of elevated blood pressure, heart rate, and body temperature since we hypothesised that these may have a different effect in older patients. Primary outcome was 30-day mortality. We extended the EWS with age and the Acutely Presenting Older Patient screener (APOP), which is a previously developed proxy for frailty. Performance of the EWS was expressed as discrimination (i.e. area under the curve (AUC)).

Results

Of 2,629 included patients, 135 (5.1%) died within 30 days. The National Early Warning Score (NEWS) performed best (AUC of 0.69). Accuracy improved up to 0.75 by adding age and subsequently the APOP screener. A deviant blood pressure, heart rate and body temperature were weaker predictors in this population than presumed in the NEWS.

Conclusions

The best performing score, NEWS, had a moderate predictive performance predicting 30-day mortality. Incorporating frailty and age in existing EWS improves predictive performance, showing the value of using frailty proxies in the ED. Assessment of acuity in the ED could also be improved by creating a geriatric EWS considering the differences in predictive strength of vital parameters in this population.

INTRODUCTION

Most hospitalised patients who deteriorate acutely have abnormal vital parameters in the hours before the event[40, 41]. Unfortunately these abnormalities or sudden changes in vital signs can be very subtle and are frequently not recognized[178]. This can result in a substantial delay of initiation of adequate treatment, and can lead to potentially avoidable detrimental outcomes. These observations resulted in the development of systems for earlier identification of deterioration and led to the introduction of early warning scores (EWS)[43]. EWS use vital signs to determine the severity of illness of a patient, and aim for earlier identification of potential severe illness with the objective to avoid the deterioration of the patient by intensifying treatment. Although EWS complements clinical decision making, vital signs can vary between specific patient populations. Therefore, cut-off values of EWS have been adjusted for paediatric patients (i.e. PEWS) and for obstetric patients (i.e. MEOWS)[45, 47]. For older patients however, such an adapted EWS does not yet exist, although studies have shown that EWS perform poor in this population[179, 180]. This may be caused by a difference in reference values of vital signs and atypical disease presentations in these older patients[181]. In addition, comorbidities and frailty are more common in older patients and are not incorporated in EWS. In triage systems however, frailty has already been of added value as a modifier, aiming to redirect early treatment to the frail patient[182].

The aim of our study was to assess the predictive performance of different EWS for 30-day mortality in older emergency department (ED) patients, and to study whether including frailty and age as predictors improved predictive performance.

METHODS

Study design and setting

We performed a secondary analysis on data of the 'Acutely Presenting Older Patient' (APOP) study, which is a prospective multi-centre cohort study in the ED of two secondary- and two tertiary-care hospitals in the Netherlands. Patients were included from September 2014 to November 2014 in Leiden University Medical Center (LUMC, Leiden), from March 2015 to June 2015 in Alrijne hospital (Alrijne, Leiderdorp), from May 2016 to July 2016 in Haaglanden Medical Center, location Bronovo (HMC Bronovo, The Hague), and from July 2016 to January 2017 in Erasmus University Medical Center (Erasmus MC, Rotterdam). In LUMC all consecutive patients who presented at the ED were included throughout the whole day. In the other participating hospitals only patients arriving between 10 AM and 10 PM could be included.

Selection of participants

All consecutive patients aged 70 years and older were eligible for inclusion. Patients were not eligible for inclusion if they did not speak Dutch nor English. Patients were approached within one hour after ED presentation. Patients in the highest triage category (i.e. immediate/red category in the Manchester Triage System[12] and patients presenting with stroke or with a ST-elevation myocardial infarction were excluded as they demand immediate treatment. Patients who were mentally unable to provide written informed consent, and of whom no proxy was available to provide informed consent were also excluded. Before approaching a patient, the treating physician or nurse was asked permission, and could refuse if they deemed the patient or family not capable of coping with the emotional or physical burden of participating. In patients with multiple ED visits, only the first visit was included. The Medical Ethics Committee of the LUMC, Alrijne Hospital, HMC Bronovo and Erasmus MC reviewed the study and deemed exempt.

Characteristics

The data collected consisted of a standardized questionnaire and an additional collection of information from the electronic patient records[183]. Data collection was performed by trained investigators. We used the Acutely Presenting Older Patient screener (APOP screener) as a proxy of frailty. This screener was developed and validated to identify patients with a high risk of either functional decline or mortality within 90 days after ED presentation[183, 184]. It consists of nine components such as age, sex, mode of arrival, self-supportiveness, recent hospital admissions and cognitive performance. In addition we retrieved data on vital signs needed for the EWS and information on admission to the ward from the electronic patient records.

Outcome

Primary outcome was mortality at 30 days after ED admission obtained from municipal records.

Early warning scores

We tested four different early warning scores, i.e. National Early Warning Score (NEWS), Modified Early Warning Score (MEWS), Rapid Emergency Medicine Score (REMS) and Rapid Acute Physiology Score (RAPS) (*Table 6.1* and *Table 6.2 a-d*).

EWS containing consciousness as a predictor were set to a lower maximum score as the AVPU and Glasgow coma scale (GCS) were not recorded[108]. Furthermore because all patients were aged seventy years or over, the REMS scored a minimum of 5 points. Thus the adjusted ranges for NEWS, MEWS, REMS and RAPS were respectively 0 to 17, 0 to 11, 5 to 22, and 0 to 12.

Table 6.1: Components within the EWS

	NEWS	MEWS	REMS	RAPS
Systolic BP	X	X		
AVPU/GCS	X	X	X	X
MAP			X	X
Heart rate	X	X	X	X
Respiratory rate	X	X	X	X
Oxygen saturation	X		X	
Body temperature	X	X		
Supplemental oxygen therapy	X			
Age			X	

Abbreviations: AVPU, alert, verbal, pain, unresponsive; BP, blood pressure; EWS, early warning score; GCS, Glasgow Coma Scale; MAP, mean arterial pressure; MEWS, modified early warning score; NEWS, national early warning score; RAPS, rapid acute physiology score; REMS, rapid emergency medicine score.

Table 6.2.a: Variables within the National Early Warning Score

	3	2	1	0	1	2	3
Body temperature, °C	≤ 35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥ 39.1	
Heart rate, bpm	≤ 40		41 - 50	51 - 90	91 - 110	111 - 130	≥ 131
Systolic blood pressure, mmHg	≤ 90	91 - 100	101 - 110	111 - 219			≥ 220
Respiratory rate, breaths/minute	≤ 8		9 - 11	12 - 20		21 - 24	≥ 25
Oxygen saturation, %	≤ 91	92 - 93	94 - 95	≥ 96			
Supplemental oxygen therapy				No		Yes	
Consciousness, AVPU	V, P, U			A			

Abbreviations: AVPU, alert, verbal, pain, unresponsive; bpm, beats per minute; ˚C, degrees Celsius; mmHg, millimetre Mercury.

Table 6.2.b: Variables within the Modified Early Warning Score

	3	2	1	0	1	2	3
Body temperature, °C		< 35.0		35.0 - 38.4		≥ 38.5	
Heart rate, bpm		≤ 40	41 - 50	51 - 100	101 - 110	111 - 129	≥ 130
Systolic blood pressure, mmHg	< 70	71 - 80	81 - 100	101 - 199		≥ 200	
Respiratory rate, breaths/minute		< 9		9 - 14	15 - 20	21 - 29	≥ 30
Consciousness, AVPU	U	P	V	A			

Abbreviations: AVPU, alert, verbal, pain, unresponsive; bpm, beats per minute; ˚C, degrees Celsius; mmHg, millimetre Mercury.

Table 6.2.c: Variables within the Rapid Emergency Medicine Score

	4	3	2	1	0	1	2	3	4	5	6
Age, years					< 45		45 - 54	55 - 64		65 - 74	> 74
Heart rate, bpm	≤ 39	40 - 54	55 - 69		70 - 109		110 - 139	140 - 179	≥ 180		
Mean arterial pressure, mmHg	≤ 49		50 - 69		70 - 109		110 - 129	130 - 159	≥ 160		
Respiratory rate, breaths/ minute	≤ 5		6 - 9	10 - 11	12 - 24	25 - 34		35 - 49	≥ 50		
Oxygen saturation, %	< 75	75 - 85		86 - 89	> 89						
GCS	≤ 4	5 - 7	8 - 10	11 - 13	≥ 14						

Abbreviations: bpm, beats per minute; GCS, Glasgow coma scale; mmHg, millimetre Mercury.

Table 6.2.d: Variables within the Rapid Acute Physiology Score

	4	3	2	1	0	1	2	3	4
Heart rate, bpm	≤ 39	40 - 54	55 - 69		70 - 109		110 - 139	140 - 179	≥ 180
Mean arterial pressure, mmHg	≤ 49		50 - 69		70 - 109		110 - 129	130 - 159	≥ 160
Respiratory rate, breaths/ minute	≤ 5		6 - 9	10 - 11	12 - 24	25 - 34		35 - 49	≥ 50
GCS	≤ 4	5 - 7	8 - 10	11 - 13	≥ 14				

Abbreviations: bpm, beats per minute; GCS, Glasgow coma scale; mmHg, millimetre Mercury.

To evaluate the effect of adding frailty and age on the predictive performance of EWS, we decided to combine the results from the APOP screener with all EWS as a dichotomous categorical variable, separating APOP high risk (i.e. > 45% 90-day mortality and functional decline) from low risk. and to add age as a continuous variable in the best performing EWS. We hypothesized that in older patient the recorded vital signs indicate a different severity of illness than in younger patients. With refitting we re-estimated the intercepts and coefficients of each predictor in the EWS[185].

Data analysis

Data were summarized using mean (standard deviation (SD)) given a normal distribution, using median (interquartile range (IQR)) in case of a skewed distribution, or using numbers with percentages [95% confidence intervals (CI)]. Patient characteristics were compared using unpaired t-tests, Mann-Whitney U test, and chi-squared test based on the distribution of the APOP high risk versus APOP low risk patients. Predictive performance of the EWS was quantified with the Area Under the Curve (AUC) with 95% CI. Subsequently, the different models were compared with the DeLong test. Missing data were replaced with multiple imputations using chained equations. In total we imputed 20 datasets.

All tests were two-sided, with a significance level of $p < 0.05$. Statistical analyses were performed using R version 3.6.1. and IBM SPSS Statistics package (version 26).

RESULTS

A total of 3,544 patients visited the ED of the four participating hospitals during the study period. 2,629 patients were included, with a predominance of females (n = 1,393, 53%) (*Figure 6.1*). Median age of the patients was 79 years (IQR: 74 - 84). The primary mode of arrival was by ambulance (50.9%) and most patients were triaged in the yellow urgency category (n = 1,534, 58.3%). Almost half of the patients were hospitalised (n = 1,289, 49.0%) (*Table 6.3*).

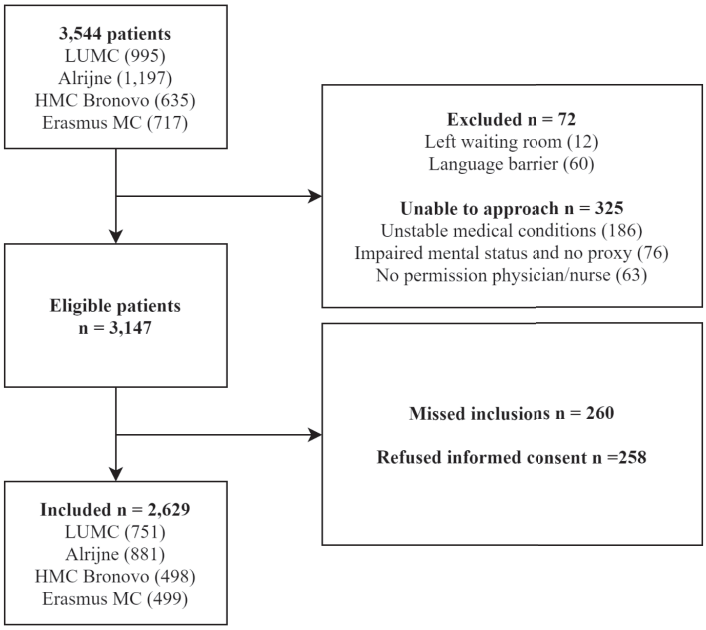


Figure 6.1: Flowchart for patient inclusion

Mortality

In the study population 135 (5.1%) patients died within 30 days of the ED visit. Patients who died were significantly older (median: 83 vs. 78), were more often men (7.0% vs. 3.4%, $p < 0.001$), had a lower systolic blood pressure (134 vs. 149 mmHg, $p < 0.001$), a lower diastolic blood pressure (74 vs. 80 mmHg, $p < 0.001$), a higher respiratory rate (22 vs. 19 per minute, $p = 0.003$), and a higher heart rate (89 vs. 83 per minute, $p = 0.006$), and had more often a APOP high risk score (46.2 vs. 18.6 % ($p < 0.001$) (*Table 6.3*).

Table 6.3: Baseline characteristics

	Total (n = 2,629)	30-day mortality (n = 135)	Alive (n = 2,494)	p-value
Demographics				
Male, n (%)	1,236 (47)	87 (64.4)	1,149 (46.1)	< 0.001
Age, median (IQR)	79 (74 - 84)	83 (77 - 89)	78 (74 - 84)	
ED presentation characteristics				
Arrival by ambulance, n (%)	1,339 (50.9)	100 (74.1)	1,239 (49.7)	< 0.001
Triage category, n (%)				
Very urgent (<10 minutes)	378 (14.4)	27 (20.0)	351 (14.1)	0.017
Supplemental oxygen therapy, n (%)	312 (11.9)	46 (37.4)	226 (13.0)	< 0.001
Vital signs at presentation				
Systolic BP in mmHg, mean (SD)	149 (28)	133.8 (27.9)	149.4 (28.0)	< 0.001
Diastolic BP in mmHg, mean (SD)	79 (17)	73.6 (16.3)	79.8 (17.1)	< 0.001
MAP in mmHg, mean (SD)	102 (18)	93.7 (18.2)	103.0 (18.0)	< 0.001
Heart rate in bpm, mean (SD)	84 (22)	88.6 (20.8)	83.4 (21.6)	0.006
Respiratory rate/min, mean (SD)	19 (6)	22.0 (8.9)	19.1 (6.4)	0.003
Oxygen saturation, median (IQR)	97 (95 - 98)	96 (94 - 98)	97 (95 - 98)	
Temperature in ˚C, mean (SD)	36.9 (0.9)	36.8 (1.1)	36.9 (0.9)	0.374
Admission, n (%)	1,289 (49.0)	105 (77.8)	1184 (47.5)	< 0.001
APOP high risk, n (%)	521 (19.8)	61 (46.2)	460 (18.6)	< 0.001

Abbreviations: APOP, acutely presenting older patient; BP, blood pressure; bpm, beats per minute; ˚C, degrees Celsius; IQR, interquartile range; mmHg, millimetre Mercury; n, number; SD, standard deviation.

The NEWS performed best in predicting 30-day mortality (AUC 0.690 (0.641 - 0.739)), and the RAPS the worst (AUC 0.522 (0.474 - 0.570)). All EWS improved when adding the result of the APOP screener as a proxy of frailty, yet the NEWS remained the best performing EWS (AUC [95% CI] 0.739 [0.691 - 0.786]). We subsequently analysed the effect of age on the predictive performance of the NEWS. This also increased the discrimination (AUC [95% CI] 0.734 [0.688 - 0.781]). Combining NEWS, the APOP screener and age improved the discriminative ability even further (AUC [95% CI] 0.752 [0.707 - 0.796]. These improvements were significant as indicated by the DeLong test ($p < 0.001$)(*Table 6.4*).

Refitting of the NEWS parameters showed that a higher blood pressure, heart rate and body temperature were weaker predictors than expressed in the NEWS (*Table 6.5*). From our data a heart rate ranging between 41 - 50 bpm even seems protective of 30-day mortality compared to the reference standard of 51 - 90 bpm.

Table 6.4: Discrimination of EWS

	AUC for 30-day mortality	95% CI
REMS	0.518	0.469 - 0.567
REMS + APOP	0.640	0.588 - 0.692
RAPS	0.522	0.474 - 0.570
RAPS + APOP high risk	0.649	0.599 - 0.699
MEWS	0.583	0.533 - 0.633
MEWS + APOP high risk	0.676	0.626 - 0.726
NEWS	0.690	0.641 - 0.739
NEWS + Age	0.734	0.688 - 0.781
NEWS + APOP high risk	0.739	0.691 - 0.786
NEWS + Age + APOP high risk	0.752	0.707 - 0.796

Abbreviations: APOP, acutely presenting older patient; AUC, area under the curve; CI, confidence interval; MEWS, modified early warning score; NEWS, national early warning score; RAPS, rapid acute physiology score; REMS, rapid emergency medicine score.

Table 6.5: Regression coefficients and SE of parameters in the NEWS

	Range of values	Weight in NEWS (rounded coefficients)	Coefficient (SE)
Systolic BP (mmHg)	≤ 90	3	2.20 (0.42)
	91 - 100	2	0.51 (0.51)
	101 - 110	1	0.81 (0.33)
	111 - 219	0	0
	≥ 220	3	-7.30 (34.10)
Heart rate (bpm)	≤ 40	3	-7.02 (37.66)
	41 - 50	1	-0.43 (1.03)
	51 - 90	0	0
	91 - 110	1	0.52 (0.22)
	111 - 130	2	-0.00 (0.36)
	≥ 131	3	0.14 (0.39
Body temperature (˚C)	≤ 35	3	0.10 (0.51)
	35.1 - 36.0	1	-0.13 (0.36
	36.1 - 38.0	0	0
	38.1 - 39.0	1	0.14 (0.29)
	≥ 39.1	3	1.19 (0.51)

Abbreviations: BP, blood pressure; bpm, beats per minute; ˚C, degrees Celsius; mmHg, millimetre Mercury; NEWS, national early warning score; SE, standard error

DISCUSSION

In this study we assessed the predictive performance of four different EWS for mortality in older ED patients. The performance of all EWS in our population was mediocre at best, with NEWS as the best performing model. When adding a proxy for frailty - i.e. high risk in the APOP screener - and age, the performance of the NEWS significantly improved.

These findings support previous research that acuity scales alone do not perform well in older patients[179, 180]. We expect that this poor performance has multiple causes. First, normality of vital signs differs, which we showed by refitting the EWS. This can be partly explained by age-related physiological changes resulting in vital parameters outside the range that is normal for a general adult population (e.g. the ability to mount high fever[186]). Also, comorbidities (e.g. hypertension) and concomitant medication can mask changes of vital signs during acute illness (e.g. beta-blockers in sepsis) and can contribute to a higher risk of undertriage[35, 87]. Therefore, potential deterioration will remain unnoticed longer. Second, frail elderly are already at risk for mortality in a normal stable situation. An acute medical deterioration leading to an ED visit has greater consequences in this subpopulation, resulting in an even higher risk of death.

Most geriatric screening tools do not use vital parameters and have neither been designed nor validated for use in the acute setting such as the ED. A systematic review concluded that solely the Identification Senior at Risk (ISAR) - the most frequently used screening tool for adverse outcomes - has a low performance in predicting adverse outcomes in the ED[187]. Geriatric screening prognosticate adverse outcomes on middle-long and long term, whereas EWS mainly predicts short term deterioration. We showed that combining both frailty and acuity can improve prediction of early mortality in the older patient population, and a potential target for early intervention during the course in the ED. This can be done by either modifying existing EWS or creating an EWS specifically for the elderly, as is done for children and pregnant women and has previously been suggested.

This study has some limitations. First, unstable and patients with a need for immediate care were excluded. These patients most likely have higher total EWS scores and the inclusion of these patients may have led to an improved risk prediction. However, in these patients a prediction tool is superfluous as it is already obvious the patient is critically ill. Second, we did not gather data on the level of consciousness, while this predictor is included in all tested EWS and apparently is considered to be an important predictor. Then again as an altered level of consciousness is such a strong predictor for mortality,

it is questionable what the additional value of applying an EWS is instead of applying treatment on the spot.

Strengths of our study are the multicentre inclusion in both secondary and tertiary referral centres and the large sample size. The inclusion of patients from several different hospitals make our findings more generalizable to other hospitals. Finally, the parameters used to execute the EWS and APOP screener can be performed right after ED arrival and is therefore feasible to use in clinical practice.

Future research in developing prediction tools should aim to include all older patients in the ED. Further studies could also shed more light on differences between the adult population and the older patient population in order to identify the true differences in presentation, triage, and vital signs. We already showed that the regression coefficients for predictors were smaller in the older population than assumed in the EWS. Presumably age above a (to be defined) cut-off point, should be added as an effect modifier or predictor in future development of EWS. Also the incorporation of a geriatric screening instrument in EWS should be evaluated. This might entail that trigger thresholds for intervention needs to be adjusted in older ED patients, and thus the introduction of a geriatric EWS is expedient.

CONCLUSION

EWS perform only moderately well in an older ED population when predicting 30-day mortality. Performance improves by adding age and a proxy for frailty such as the APOP screener. Also, the predictive strength of vital parameters is different in older patients. Assessment of acuity in the ED could be improved by incorporating age and frailty as new parameters existing EWS, or by creating an elderly specific EWS.

chapter 7

Predicting 30-day Mortality Using Point-of-care testing; an External Validation and Derivation Study

Anniek Brink
Romy Schuttevaer
Jelmer Alsma
Robert Zietse
Stephanie C.E. Schuit
Hester F. Lingsma

PLoS One. 2020 Sep 24;15(9):e0239318.

ABSTRACT

Background

Early risk stratification for guiding treatment priority in the emergency department (ED) is becoming increasingly important. Existing prediction models typically use demographics, vital signs and laboratory parameters. Laboratory-based models require blood testing, which may cause substantial delay. However, these delays can be prevented by the use of point-of-care testing (POCT), where results are readily available. We aimed to externally validate a laboratory-based model for mortality and subsequently assessed whether a POCT model yields comparable performance.

Methods

All adult patients visiting the ED of a university hospital between January 1st, 2012 and December 31st, 2016 were retrospectively reviewed for inclusion. Primary outcome was defined as 30-day mortality after ED presentation. We externally validated one existing prediction model including age, glucose, urea, sodium, haemoglobin, platelet count and white blood cell count. We assessed the predictive performance by discrimination, expressed as Area under the Curve (AUC). We compared the existing model to an equivalent model using predictors that are available with POCT (i.e. glucose, urea, sodium and haemoglobin). Additionally, we internally validated these models with bootstrapping.

Results

We included 34,437 patients of whom 1,942 (5.6%) died within 30 days. The AUC of the laboratory-based model was 0.794. We refitted this model to our ED population and found an AUC of 0.812, which decreased only slightly to 0.790 with only POCT parameters.

Conclusions

Our POCT-model performs similar to existing laboratory-based models in identifying patients at high risk for mortality, with results available within minutes. Although the model needs further validation and evaluation, it shows the potential of POCT for early risk stratification in the ED.

Keywords

Prediction, Emergency Department, Mortality, Laboratory parameters

INTRODUCTION

Identifying patients in the emergency department (ED) at risk of dying remains challenging. The existing prediction models are typically based on demographics and vital signs.

Triage systems are initially used to identify the most severely ill patients. However, current triage systems, such as the Emergency Severity Index (ESI)[188] and the Manchester Triage System (MTS)[12], were mainly introduced for trauma patients. The performance of triage systems in all ED patients is poor[127, 189, 190]. Early warning scores (EWS) are also used in the ED, either as replacement or as addition to triage[191, 192]. Examples of EWS are the Modified Early Warning Score (MEWS) and the National Early Warning Score (NEWS). However, EWS are developed to detect inpatient clinical deterioration and the predictive value of EWS for mortality in the ED varies[86, 193] and both validation and calibration are inadequate or lacking[194].

Using laboratory parameters is another approach for early risk stratification. An advantage of laboratory parameters in prediction models is their objective measurement. *Asadollahi et al.* provided a laboratory-based prediction model derived from 1,650 acute medical and surgical patients, which performed well with an AUC of 0.848[195]. The model was internally-externally validated using data from the same hospital in a different period of time. The model uses age, glucose, urea, sodium, haemoglobin, platelets and white blood cell count as predictors. These six laboratory parameters were selected from a large array of potential parameters and are known to correlate with adverse outcome[196-199]. Laboratory models require blood testing and can therefore cause a substantial delay if blood samples are analysed by a central laboratory. However, these delays might be prevented by point-of-care testing (POCT), which yields results within minutes[200, 201].

The aim of this study is to determine whether a laboratory model can be implemented using only POCT laboratory testing. Since external validation is a critical step to implementation in clinical practice, and to potentially improve the feasibility of the model, the first aim of this study is to externally validate the laboratory-based model by *Asadollahi et al.*[195] in a large unselected population of ED patients. The second aim is to assess whether a model based only on POCT available laboratory parameters yields comparable performance.

METHODS

Study design and setting

We performed a retrospective cohort study in the ED of the Erasmus University Medical Center Rotterdam (Erasmus MC), Rotterdam, the Netherlands, which is a large tertiary referral centre, situated in an urban area. The ED has approximately 32,000 visits annually. Data from all patients were automatically extracted from the electronic health records on a regular basis and collected in a database.

Selection of participants

We included all ED visits from January 1st 2012 up to December 31st 2016. Adult patients, aged 18 years and over, in which laboratory diagnostics were performed were selected for this study. Per patient only the first visit to the ED was included.

Measurements and outcomes

We extracted demographic data (i.e. sex, age) and presenting vital parameters (i.e. body temperature, heart rate, respiratory rate, oxygen saturation, blood pressure and consciousness level using AVPU-scale). Furthermore, we extracted acuity scale according to MTS category, disposition (i.e. in hospital admission), and arrival (i.e. by ambulance). In line with the study of *Asadollahi et al.* we extracted haemoglobin, serum sodium, plasma glucose, white blood cell count, serum urea and platelet count. These laboratory values were afterwards categorized (*Table 7.1*).

Haemoglobin levels were converted from mmol/L to g/dL. Subsequently, we selected the laboratory parameters that are measurable with POCT, i.e. glucose, urea, sodium and haemoglobin. At Erasmus MC, the ABL800 FLEX (Radiometer America Inc., Westlake, Ohio) blood gas analyser for POCT is used, which yields results within two minutes. Outcome was defined as 30-day mortality after the index ED visit. Mortality data were retrieved from the patient records, which are linked to municipal mortality records. The Medical Ethics Committee of the Erasmus MC reviewed the study and concluded that our study did not fall under the scope of the Medical Research Involving Human Subjects Act and therefore no informed consent needed to be obtained.

Statistical analysis

Patient characteristics were presented as mean (SD), median (interquartile range (IQR)) or absolute numbers (percentage), when appropriate. Missing data were handled using multiple imputations (n = 5) with a chained equations procedure, which means that the expected value of the missing data point is estimated based on the available data.

We examined all patient characteristics of patients that were alive versus the patients who died within 30 days from the index ED visit. Data were compared using Pearson chi-squared tests or unpaired t-tests, based on distribution of data.

Model performance of the laboratory-based model was described as discrimination and calibration. Discrimination was assessed using the area under the Receiver Operating Characteristic-curve (AUC). We assessed calibration with a calibration plot in which the slope indicates the relation between the observed and the predicted outcome (i.e. ideally close to 1) and the intercept indicates whether the predictions are systematically deviant (i.e. ideally close to 0). We calculated likelihood ratios for all cut-off points from the total score and determined the ideal cut-off point using Youden's index (i.e. the cut-off point combining the optimal sensitivity and specificity). Interval likelihood ratios were established for several different intervals. Furthermore, we refitted the model on our data and subsequently reduced the model by only including parameters which could be tested using POCT. These models were internally validated using five hundred times bootstrap resampling.

Analyses were conducted with IBM SPSS Statistics for Windows version 26 (IBM Corp., Armonk, New York, USA) and R statistics version 3.6.1. A significance level of $p < 0.05$ was considered as statistically significant.

RESULTS

Patient characteristics

116,398 adult ED visits were recorded between January 2012 and December 2016. Laboratory testing was performed in 54,753 of these visits. Selecting only first ED visits, yielded 34,437 patients eligible for analysis (*Figure 7.1*). The majority of the population was male (54.8%). Median age (IQR) was 54 years (37 - 67). Admission rate was 55.7% and in total 1,942 (5.6%) patients died within 30 days after the ED visit. Patients who died presented to the ED with more abnormal vital signs (i.e. higher heart rate (93 vs. 87 per minute), lower systolic blood pressure (135 vs. 140 mmHg), abnormal consciousness level (37.3 vs. 7.5%), $p < 0.001$) and were significantly older (68 vs. 52 years, $p < 0.001$) (*Table 7.2*).

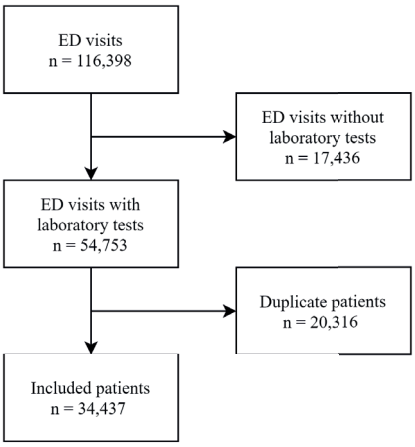


Figure 7.1: Flowchart of patient inclusion

Model performance

Haemoglobin was most frequently tested (in 98.8% patients), whereas blood urea nitrogen was least often tested (72.0%). All predictive effects we found corresponded to the original model (e.g. low platelet count, haemoglobin level, sodium levels were associated with 30-day mortality)(*Table 7.1* and *7.2*). Of all predictors in the model, age ≥ 65 years was the strongest predictor for 30-day mortality in univariate analysis (OR [95% CI] = 4.4 [4.0 - 4.8]) (*Table 7.3*).

Table 7.1: Laboratory tests with its cut-off points

Parameter (unit)	Reference range
Age ≥ 65 (years)	
Urea > 7.0 (mmol/L)	2.5-7.5 (mmol/L)
Haemoglobin < 12.0 (g/dL) (7.45 mmol/L)	♂: 14 - 17.5 (g/dL) ♀: 12.3 - 15.3 (g/dL)
Sodium < 135 (mmol/L)	135 - 145 (mmol/L)
Glucose > 7.0 (mmol/L)	< 7.8 (mmol/L)
White blood count > 10.0 (*10^9/L)	4.0 - 10.0 (*10^9/L)
Platelet count < 150 (*10^9/L)	150 - 400 (*10^9/L)

Table 7.2: Baseline characteristics

	All patients	Died within 30-days	Alive	p-value
N (%)	34,437 (100%)	1,942 (5.6)	32,495 (94.4)	
Demographics				
Male, N (%)	18,827 (54.7)	1,187 (61.1)	17,640 (54.3)	< 0.001
Age, median (IQR)	54 (37-67)	68 (58-78)	52 (36-66)	
ED presentation characteristics				
Arrival by ambulance, N (%)	10,387 (30.2)	1,074 (55.3)	9,313 (28.7)	< 0.001
Triage category, N (%)[a], Immediate / very urgent	6,827 (19.8)	1,015 (52.3)	5,812 (17.9)	< 0.001
Vital signs				
Body temperature in °C[b], mean (SD)	36.8 (0.98)	36.2 (1.5)	36.9 (0.9)	< 0.001
Heart rate in bpm[c], mean (SD)	87 (21)	93 (25)	87 (21)	< 0.001
Systolic blood pressure in mmHg[d], mean (SD)	139 (27)	135 (39)	140 (26)	< 0.001
Diastolic blood pressure in mmHg[e], mean (SD)	82 (17)	77 (24)	82 (16)	< 0.001
Respiratory rate per minute[f], mean (SD)	19 (7)	21 (8)	19 (7)	< 0.001
Oxygen saturation in %[g], median (IQR)	98 (96-99)	97 (94-99)	98 (96-99)	< 0.001
Consciousness not alert[h], N (%)[b]	3,148 (9.1)	712 (37.3)	2,436 (7.5)	< 0.001
Admission, N (%)	19,172 (55.7)	1,607 (84.2)	17,565 (54.0)	< 0.001
Laboratory tests				
Urea (mmol/L)[i], mean (SD)	6.7 (5.1)	9.7 (7.6)	6.5 (4.9)	< 0.001
Sodium (mmol/L)[j], mean (SD)	139 (4.4)	138 (6.1)	139 (4.2)	< 0.001
Glucose (mmol/L)[k], mean (SD)	7.4 (3.8)	8.9 (5.6)	7.2 (3.5)	< 0.001
Haemoglobin (g/dL)[l], mean (SD)	13.3 (2.1)	12.3 (2.6)	13.3 (2.1)	< 0.001
White blood cell count (10^9/L)[m], mean (SD)	10.3 (18.3)	13.8 (14.6)	10.1 (18.4)	< 0.001
Platelets (10^9/L)[n], mean (SD)	256 (104)	242 (133)	257 (243)	< 0.001

Missing data are not yet imputed.
Abbreviations: °C , degrees Celsius; bpm, beats per minute; dL, decilitre; ED, emergency department; g, gram; IQR, interquartile range; L, litre; mmol, millimole; N, number; SD, standard deviation.
[a]Data on triage category were missing for 2,610 (7.6%) patients.
[b]Data on body temperature were missing for 10,995 (31.9%) patients.
[c]Data on heart rate were missing for 5,333 (15.5%) patients.
[d]Data on systolic blood pressure were missing for 5,499 (16.0%) patients.
[e]Data on diastolic blood pressure were missing for 5,452 (15.8%) patients.
[f]Data on respiratory rate were missing for 17,321 (50.3%) patients.
[g]Data on oxygen saturation were missing for 6,355 (18.5%) patients.
[h]Data on conscious level were missing for 9,837 (28.6%) patients.
[i]Data on urea level were missing for 9,640 (28.0%) patients.
[j]Data on sodium level were missing for 1,108 (3.2%) patients.
[k]Data on glucose level were missing for 955 (2.8%) patients.
[l]Data on haemoglobin level were missing for 421 (1.2%) patients.
[m]Data on white blood cell count were missing for 2,229 (6.5%) patients.
[n]Data on platelet count were missing for 1,786 (5.2%) patients.

External validation showed an AUC [95% CI] of 0.796 [0.788 - 0.806] (*Figure 7.2*). The calibration curve had a slope of 0.77 and an intercept of 0.34 (*Figure 7.3*).

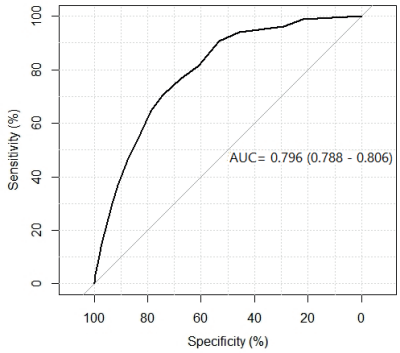


Figure 7.2: Performance of the laboratory model

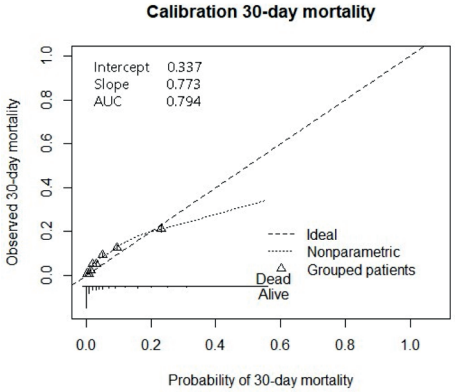


Figure 7.3: Calibration curve of the laboratory model

We refitted the laboratory-based model on our own dataset with subsequent reduction to a model with only POCT parameters (*Table 7.3*). The refitted laboratory-based model yielded an internally validated AUC of 0.813, which slightly decreased to 0.790 when only including age and POCT parameters.

Likelihood ratios of score intervals of 0 to 5, 6 to 13 and 14 to 20 were 0.31, 1.77 and 5.04 respectively. Positive and negative likelihood ratios for dichotomous cut-off points are found in *Table 7.4*. The highest Youden's index was found using a cut-off point score of 8.

Table 7.3: Odds ratios [95% CI] for the full model and POCT model

Parameter	Odds ratio [95% CI]	Odds ratio [95% CI]	Odds ratio [95% CI]
	Univariate	Full model	POCT model
Age ≥ 65 (years)	4.21 [3.83 - 4.63]	2.73 [2.45 - 3.05]	2.60 [2.34 - 2.89]
Urea > 7.0 (mmol/L)	3.24 [2.85 - 3.67]	1.61 [1.39 - 1.88]	1.71 [1.48 - 1.98]
Haemoglobin < 12.0 (g/dL)	2.57 [2.34 - 2.83]	1.73 [1.56 - 1.92]	1.82 [1.65 - 2.02]
Sodium < 135 (mmol/L)	2.63 [2.36 - 2.96]	1.52 [1.35 - 1.72]	1.64 [1.45 - 1.85]
Glucose > 7.0 (mmol/L)	4.36 [3.95 - 4.81]	2.82 [2.54 - 3.13]	3.15 [2.85 - 3.50]
White blood cell count > 10.0 (*10^9/L)	2.39 [2.15 - 2.66]	2.29 [2.05 - 2.57]	NA
Platelet count < 150 (*10^9/L)	2.89 [2.57 - 3.25]	2.79 [2.44 - 3.19]	NA

Abbreviations: CI, confidence interval; dL, decilitre; g, gram; mmol, millimole; L, litre; NA, not applicable. The linear predictor of the full model can be calculated with the following formula: LP = -4.790 + 1.005*(Age ≥ 65) + 0.479*(Urea > 7.0) + 0.548*(Haemoglobin < 12.0) + 0.420*(Sodium < 135) + 1.037*(Glucose > 7.0) + 0.830*(White blood count > 10.0) + 1.025*(Platelet count < 150). The linear predictor of the POCT model: LP = -4.302 + 0.995*(Age ≥ 65) + 0.539*(Urea > 7.0) + 0.601*(Haemoglobin < 12.0) + 0.494*(Sodium < 135) + 1.149*(Glucose > 7.0). To determine the individual risk on 30-day mortality, apply the following formula: 1/(1 + exp(–linear predictor).

Table 7.4: Sensitivity, specificity, LR+, LR-, Youden's index for different cut-off points

Cut-off point	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Youden's index
1	0,99	0,22	1,27	0,05	0,21
2	0,99	0,22	1,27	0,05	0,21
3	0,96	0,29	1,36	0,13	0,25
4	0,94	0,46	1,73	0,13	0,40
5	0,91	0,53	1,94	0,18	0,44
6	0,81	0,61	2,08	0,31	0,42
7	0,77	0,68	2,38	0,34	0,44
8	0,70	0,74	2,74	0,40	0,45
9	0,65	0,79	3,03	0,45	0,43
10	0,53	0,84	3,34	0,56	0,37
11	0,46	0,88	3,74	0,61	0,34
12	0,37	0,91	4,25	0,69	0,28
13	0,30	0,93	4,46	0,75	0,23
14	0,19	0,96	5,04	0,84	0,15
15	0,14	0,98	5,51	0,89	0,11
16	0,09	0,98	5,32	0,93	0,07
17	0,04	0,99	6,94	0,97	0,03
18	0,03	1,00	7,38	0,97	0,03
19	0,00	1,00	33,47	1,00	0,00
20	0,00	1,00	NA	1,00	0,00

Interval likelihood ratios for score 0 to 5, 6 to 13 and 14 to 20 were 0.31, 1.77 and 5.04. Positive likelihood ratio = sensitivity/(1-specificity). Negative likelihood ratio = (1-sensitivity)/specificity.

DISCUSSION

In this study we first externally validated the model by *Asadollahi et al.* that uses laboratory parameters and the patients' age to predict mortality. Next we assessed that the *Asadollahi et al.* model based only on POCT available laboratory parameters yielded comparable performance. To our knowledge, we are the first to validate this laboratory-based model and to perform calibration. Our external validation resulted in a reasonable AUC of 0.796. One benefit of our study is that we used a large database for the external validation, which limits uncertainty in the performance of the model and thus increases the clinical relevance[202].

Despite the fact that most models are introduced without calibration, it is a critical step preceding implementation of a model. Calibration describes the agreement between the calculated, based on the prediction model, and the observed number of occurrences. The calibration of this model was suboptimal, which indicates the model slightly overestimates the mortality risk.

A major disadvantage of the study by *Asadollahi et al.* is the case-controlled study design. The authors included deceased and non-deceased patients in a 1:2 ratio, which yields a mortality rate of 33%. This results in an overestimation of the prevalence of 30-day mortality. Since we conducted a cohort study, our mortality rate reflects the 30-day mortality prevalence more accurately.

One of the merits of the model by *Asadollahi et al.* is that it uses only a few parameters. Additionally, the parameters within this model are virtually always assessed in patients with indication for laboratory diagnostics admitted to the ED. Therefore, this model is generally applicable and easy to use by clinicians. Its simple interpretation accommodates usage implementation in electronic patient files. The power of laboratory values in prediction research is that they provide an objective measurement, especially compared to manually collected vital parameters. Vital parameters that are manually collected are prone to interrater variability. Also, vital parameters are subject to influences that are not always taken into account (pain, stress, normality for an individual patient). A downside of laboratory values is they take time to become available and are therefore difficult to implement in a decision model in the ED.

Ideally, prediction models in the ED should consist of readily available parameters and as little parameters as possible, making the model convenient for clinical practice. As most laboratory test results take more than an hour, the second aim of our study was to assess whether a model based on POCT parameters yielded similar predictive per-

formance in predicting 30-day mortality. We found that our model with age and POCT parameters had similar performance. We provided the regression coefficients and an intercept allowing to replicate this study, but also to facilitate implementation of this model in clinical practice. POCT is promising since it only takes minutes to analyse blood samples. This may lead to a reduction in time to diagnosis and initiation of treatment. Furthermore, POCT-systems are already used in EDs. Although presently not every ED has a POCT-analyser, which may limit the applicability of this study, this study may encourage to invest in POCT-analysers.

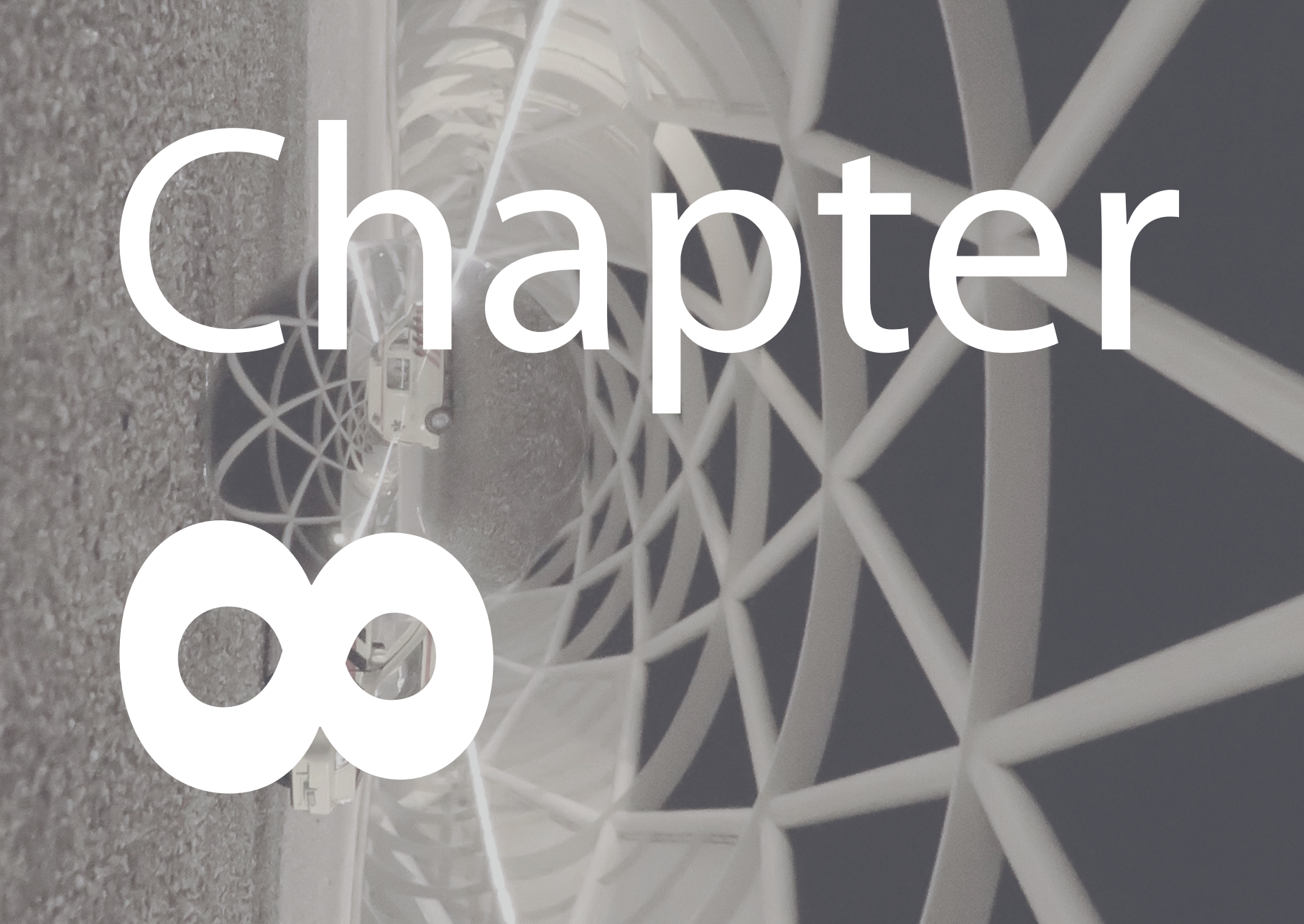
A general limitation of literature concerning prediction models in the ED, is that hardly any study provides sufficient information to execute external validation. There are several models based on laboratory values published which performed well in general with an AUC above 0.80[133, 137, 138].

A limitation of our study is its retrospective study design, which makes our study prone to bias. Nevertheless, laboratory data were automatically retrieved from the laboratory testing machines thus the quality of the data is high and not subject to human mistakes. Furthermore, we had missing data, mainly vital signs, which we replaced using multiple imputation. This is a valid way to manage even large samples of missing data[162], although a database with all data available is obviously superior. Therefore, we should strive to collect data as complete as possible. Last, this validation study took place in a tertiary care centre which corresponds to the derivation study[195]. Therefore, our results might be less generalizable to other centres with patients with different complexity and pathology. We therefore recommend external validation of our model in another centre, before implementation. In addition, we encourage considering POCT in prediction model development, researching both its discrimination and calibration.

In conclusion, the performance of the model by *Asadollahi et al.* was adequate in identifying patients at high risk for mortality in the ED. However, our POCT-model performs similar with results available within minutes. Although our model needs further validation and evaluation, it shows the potential of POCT in early risk stratification in the ED.

Section III

Drug Prescription and Adherence in the Emergency Department


Chapter
8

Non-adherence to Antimicrobial Guidelines in Patients with Bloodstream Infection Visiting the Emergency Department

Romy Schuttevaer
Anniek Brink
Jelmer Alsma
Willian van Dijk
Damian C. Melles
Jurriaan E.M. de Steenwinkel
Hester F. Lingsma
Annelies Verbon
Stephanie C.E. Schuit

Eur J Intern Med. 2020 Aug;78:69-75.

ABSTRACT

Background

Non-adherence to antimicrobial guidelines in patients with bloodstream infection can be considered as overtreatment, undertreatment or equivalent treatment, and could lead to suboptimal care. Our aim was to examine the association between non-adherence and appropriate coverage as well as to assess the impact of non-adherence on 30-day mortality.

Methods

We conducted a retrospective cohort study between 2012 and 2017 at a tertiary university hospital. All adult patients attending the emergency department with a bloodstream infection were included. Adherence was defined as guideline-recommended antibiotic therapy. Non-adherence was either undertreatment (too narrow-spectrum), overtreatment (too broad-spectrum), or equivalent treatment. Primary outcomes were appropriate coverage (i.e. antibiotic therapy that matches *in vitro* susceptibility of the isolated bacteria) and 30-day mortality.

Results

We included 909 patients of whom only 395 (43.5%) were treated adherently, 72 (7.9%) received an equivalent treatment, 87 (9.6%) were overtreated, and 355 (39.1%) were undertreated. Overtreated patients were more severely ill, whilst undertreated patients had more favourable patient characteristics. Overtreatment did not result in higher appropriate coverage, whereas undertreatment was associated with lower coverage (OR [95%CI]: 0.18 [0.12 - 0.26]). Overtreatment and undertreatment were not associated with 30-day mortality.

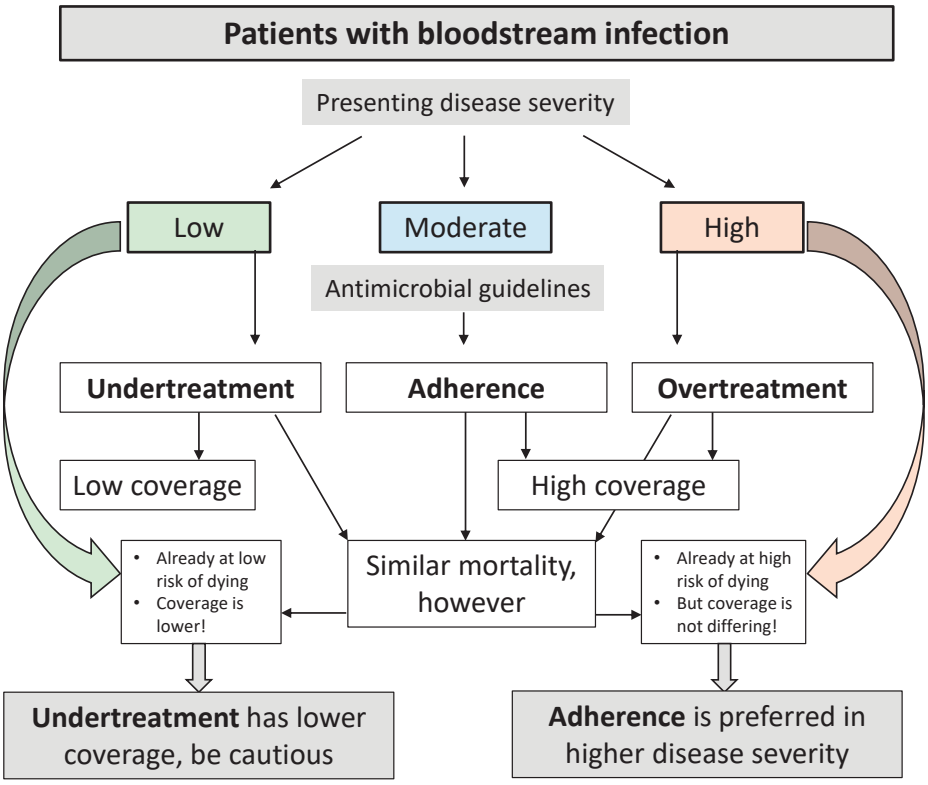
Conclusions

Guideline adherence likely depends on disease severity, because overtreatment was more often observed in patients with high disease severity and undertreatment in less severely ill patients. Undertreatment had no survival disadvantages in less severely ill patients, however, appropriate coverage was significantly lower. Overtreatment was neither associated with higher appropriate coverage nor a survival benefit compared to adherence. Therefore, overtreatment seems not justifiable.

Keywords

Antimicrobial guidelines, Adherence, Bloodstream infection, Emergency department.

GRAPHICAL ABSTRACT



INTRODUCTION

Bacterial infections can result in considerable mortality and have a profound global burden[203-205]. Patients with a severe infection (e.g. sepsis) often present in an acute care setting, such as the emergency department (ED). To provide proper care in this setting, initiation of the antibiotic therapy that matches in vitro susceptibility of the causative bacteria (i.e. with appropriate coverage) is important[206]. However, the causative pathogen has yet to be identified by cultures and this process usually takes over 24 hours. Therefore, antibiotic therapy in the ED is virtually always initiated empirically[206].

For patients with a suspected bacterial infection, guideline recommendations for empiric antibiotic therapy should depend on local prevalence of pathogens and antimicrobial resistance patterns[207]. Such antimicrobial guidelines usually provide recommendations for a specific working diagnosis (i.e. suspected source of infection). The aim of antimicrobial guidelines is to ensure that the antibiotic therapy with appropriate coverage is given before culture results become available, thereby preventing mortality. In addition, guidelines aim to reduce misuse of broad-spectrum antibiotic therapy, in order to prevent selection of antimicrobial resistance and adverse effects[208].

Non-adherence to antimicrobial guidelines in patients with a proven bloodstream infection (BSI) is disadvantageous when it results in inappropriate coverage[209, 210]. However, literature about non-adherence in the ED is scarce and discrepant. Rate of non-adherence ranged from 10 to 53% and these studies did not differentiate between different types of non-adherence (i.e. equivalent, over-, or, undertreatment)[211-213]. Therefore, we intended to evaluate non-adherence to antimicrobial guidelines for adult patients with BSI attending the ED. Our aims were: primarily to examine the association between the different types of non-adherence and appropriate coverage, and secondly to assess the impact of the different types of non-adherence on 30-day mortality.

MATERIALS AND METHODS

Study design and setting

We conducted a retrospective cohort study at the Erasmus University Medical Center Rotterdam (Erasmus MC), which is a tertiary university hospital in the Netherlands. We used data from all patients attending the ED with BSI from July 2012 through December 2017.

Selection of participants

Patients were eligible for inclusion if they were at least 18 years of age, had a guideline-specified working diagnosis, and had a laboratory proven bacterial BSI in the ED. BSI was defined as presence of a known pathogen (e.g. E. coli) in one blood culture or a less pathogenic bacteria (e.g. S. epidermidis) in at least two blood cultures collected on separate occasions within two days from ED admission[214, 215]. Only the first episode of BSI was included to prevent domination of results by individuals that frequently visited the ED.-

Data collection and processing

Data were derived from an ED database and combined with a database from the Medical Microbiology, containing all collected blood cultures. The ED database included the working diagnosis, empiric antibiotic therapy administered during the ED visit, other patient characteristics, and mortality. General and demographic presenting patients characteristics collected were: sex, age, arrival (by ambulance or not), triage category (according to the Manchester Triage System)[216], disposition (direct intensive care unit admittance or other), chills[217], vomiting[217], need for vasopressors, and origin of infection (nosocomial or community-acquired)[218]. To further account for initial severity of disease we used the first recorded vital signs (i.e. body temperature, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, and consciousness), whether there was need for any supplemental oxygen, and calculated the National Early Warning Score (NEWS)[219, 220] (*Appendix 8.A*). Additionally, to account for comorbidity we collected all components of the age-adjusted Charlson Comorbidity Index (CCI)[221] (*Appendix 8.B*). For mortality data we used municipal death registration records.

Patients with a positive blood culture in the ED were identified via the positive blood culture database of Medical Microbiology. This database contained information about type of pathogen and the susceptibility (antibiogram). Blood cultures were performed using the BACTEC system (Becton Dickinson Diagnostic Instrument Systems, Sparks, Md) according to the manufacturers protocol. Type of pathogen was identified by MALDI-TOF MS analysis (Microflex, Bruker Daltonics, Bremen, Germany). The in vitro susceptibility testing was performed using the VITEK 2 (bioMérieux, Marcy l'Etoile, France) system. Based on earlier applied antibiotic therapy in the ED and established in vitro susceptibility of the isolated pathogen, we determined whether coverage of the empiric therapy was appropriate or not. In accordance with previous studies, no empiric antibiotic therapy, ineffective antibiotic therapy (based on antibiogram or dosage) or not intravenously administered antibiotic therapy (except for antibiotics with high bioavailability such as Ciprofloxacin and Metronidazole) were all considered as inappropriate coverage of the isolated pathogen[59].

Adherence to guidelines was defined as initial antibiotic therapy administered in the ED in accordance with local hospital guideline recommendations for empiric antibiotic therapy. This definition corresponds to previous definitions of adherence in comparable study settings[211-213]. Our empiric guideline recommendations depend on national antimicrobial guidelines and are updated based on local prevalence and resistance patterns[222]. Guidelines provide recommendations for a specific working diagnosis, and are easily available online for all physicians in our hospital[223]. Guideline deviation was considered adherent if a proper motivation was described in the medical chart, i.e. if altered based on previous relevant cultures (only to broaden therapy), renal function (e.g. applying an alternative to Gentamicin while preserving the antimicrobial spectrum if pre-existent glomerular filtration rate was < 30 millilitre/minute or after recent renal transplantation without proper transplant function), and comorbidity (e.g. sickle cell disease, functional asplenia). Additionally, empiric antibiotic therapy was considered adherent if altered after direct consultation with a clinical microbiologist or infectious diseases specialist. In case of multiple working diagnoses, all highly suspected diagnoses needed to be covered. Absence of antibiotic prescription was considered adherent in case of a suspected cholecystitis (if not severely ill and if not immunocompromised) and gastro-enteritis (if not recently returned from traveling, without (persisting) high fever, no dysentery, and if not immunocompromised).

Conversely, non-adherence was defined as failure to treat in accordance with our hospital guideline. Previous studies did not divide non-adherence into an equivalent, overtreatment, and undertreatment group[211-213]. We scored non-adherence as undertreatment if therapy was more narrow-spectrum than guideline-recommended therapy (e.g. not administering antibiotics, not administering recommended Gentamicin). Overtreatment was scored if antibiotic therapy was more broad-spectrum than guideline-recommended therapy (e.g. administering additional antibiotic agents while not recommended). If antibiotic therapy was non-adherent, but equivalent with regard to spectrum, a separate equivalent group was introduced (e.g. Amoxicillin/clavulanic acid with Gentamicin is equivalent to Cefuroxime with Gentamicin for cholangitis, unknown sepsis, and urosepsis). Equivalent treatment was either in accordance with national antimicrobial guidelines, or not. For a detailed description of non-adherence scoring, see *Appendix 8.C*.

Three authors (RS, AB, JA) independently reviewed medical charts to score both working diagnoses and adherence. Scoring was discussed during expert meetings with clinical microbiologists, infectious diseases specialists, and acute physicians.

Data analysis

We examined all presenting patient characteristics that reflect severity of disease among adherently versus (vs.) non-adherently (i.e. equivalent, over-, under-) treated patients. Based on distribution, data were compared using unpaired t-tests, chi-squared tests, or Fisher's exact tests. Distribution of these patient characteristics will reveal whether there are differences in initial disease severity between adherently, equivalently, overtreated, and undertreated patients.

First, we conducted inferential statistics to investigate the association between non-adherence and appropriate antibiotic coverage with univariable logistic regression. We did not control for patient characteristics because we assume they affect appropriate antibiotic coverage only through (non)-adherence. However, secondly, for the association between non-adherence and 30-day mortality we did expect confounding by patient characteristics and therefore we used multivariable logistic regression to limit bias. We considered patient characteristics as confounders during further analyses if, based on expert knowledge, the characteristic was associated with (non)-adherence and 30-day mortality[224]. Additionally, we repeated the analyses for undertreatment after excluding patients that received no antibiotic therapy.

Results were presented as odds ratios (OR) with 95% confidence intervals (CI). All hypothesis tests were 2-sided, with a significance level of $p < .05$. We handled missing data using multiple imputations. For efficiency purposes we imputed 20 datasets using the chained equations method. Statistical analyses were performed using R version 3.4.4.

RESULTS

Patient characteristics

We identified 1.286 adult patients with a positive laboratory proven blood culture taken in the ED. We excluded 247 patients with a recurrent BSI during our study period, resulting in 1.039 unique patients with BSI (*Figure 8.1*). 909 patients had a guideline-specified working diagnosis, which are shown in *Table 8.1*. Most prevalent working diagnoses among patients with BSI were urosepsis or pyelonephritis (n = 266, 29.3%) and cholangitis (n = 181, 19.9%). In 893 (98.2%) patients we found a known pathogen (e.g. 311 *Escherichia coli*) and in 16 (1.8%) we found a common commensal on multiple blood cultures collected on separate occasions within two days from ED admission (e.g. 11 Staphylococcus epidermidis)(*Appendix 8.D*). 30-day mortality was 11.4%.

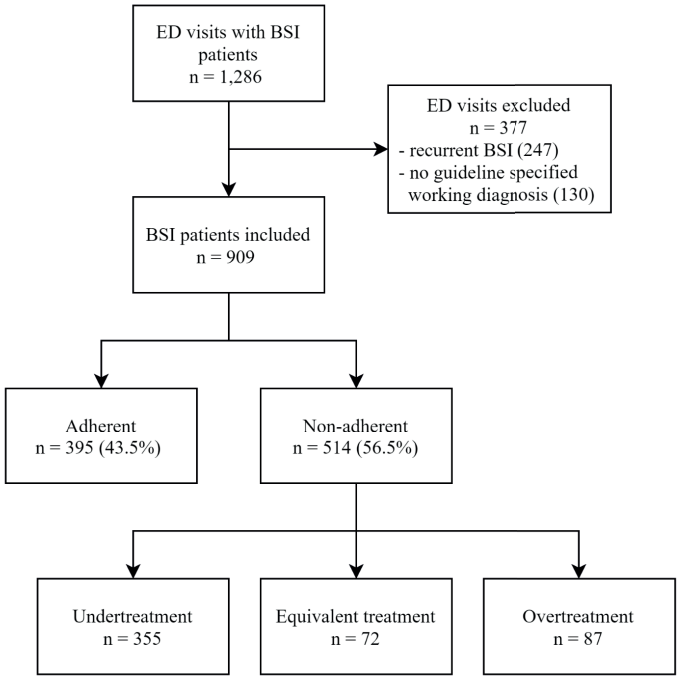


Figure 8.1: Flowchart of study selection.
Abbreviations: ED, emergency department; BSI, bloodstream infection

Table 8.1: Working diagnoses and guideline-recommended antibiotic therapy in patients with bloodstream infection at the emergency department

Suspected infection focus	Working diagnosis	N (%)	Subcategory	Guideline-recommend antibiotic therapy
Unknown	Sepsis	98 (10.8)	CA HA	Cefuroxime and Gentamicin Piperacillin/tazobactam and Gentamicin
Febrile neutropenia	Sepsis	37 (4.1)		Meropenem
Urogenital	Sepsis or pyelonephritis	266 (29.3)	CA HA	Cefuroxime and Gentamicin Piperacillin/tazobactam and Gentamicin
Respiratory	Mild pneumonia (CURB 0-1)	45 (5.0)	CA	Amoxicillin[1]
	Moderate pneumonia (CURB 2)	26 (2.9)	CA	Amoxicillin
	Severe pneumonia (CURB 3-5)	37 (4.1)	CA	Amoxicillin/clavulanic acid and Ciprofloxacin[1]
	Pneumonia	25 (2.8)	HA	Piperacillin/tazobactam (and Gentamicin if doubt about source or if septic)

Table 8.1: Working diagnoses and guideline-recommended antibiotic therapy in patients with bloodstream infection at the emergency department (*continued*)

Suspected infection focus	Working diagnosis	N (%)	Subcategory	Guideline-recommend antibiotic therapy
	Aspiration	9 (1.0)	CA	- Amoxicillin/clavulanic acid[1] - Cefuroxime and Metronidazole[1]
	Pulmonic abscess /pleura empyema	9 (1.0)		Amoxicillin/clavulanic acid
Abdominal	Sepsis	29 (3.2)	CA HA	Cefuroxime and Metronidazole and Gentamicin Piperacillin/tazobactam and Gentamicin
	Cholangitis	181 (19.9)		Cefuroxime and Gentamicin
	Peritonitis, primary (SBP)	13 (1.4)		Ceftriaxone
	Peritonitis, secondary	11 (1.2)	CA	Cefuroxime and Metronidazole[2] and Gentamicin
	Gastro-enteritis	18 (2.0)	CA, returned from traveling; if (persisting) high fever, dysentery, immunocompromised	Initially without antibiotic therapy - Azithromycin[1] - Erythromycin and Ciprofloxacin
Skin, soft tissue, bone	Cellulitis	22 (2.4)		Flucloxacillin[1]
	Erysipelas	14 (1.5)		Penicillin[1]
Central nervous system	Meningitis, primary	32 (3.5)	Before 2015: < 50 years, not immunocompromised	Ceftriaxone and Amoxicillin Ceftriaxone
Intravascular, thorax	Intravascular catheter	19 (2.1)		- Vancomycin - Cefuroxime and Gentamicin

Only working diagnoses with a prevalence ≥ 1.0% are shown in this table.
Abbreviations: CA, community-acquired; HA, hospital-acquired; CURB65, confusion, blood urea nitrogen, respiratory rate, systolic blood pressure, age ≥ 65; SBP, spontaneous bacterial peritonitis.
All antibiotic therapy had to be administered intravenously, except for [1]oral and [2]intraperitoneal administration was allowed.

Treatment was adherent for 395 (43.5%) patients, 72 (7.9%) received equivalent treatment, 87 (9.6%) were overtreated, and 355 (39.1%) were undertreated. Equivalently treated patients had therapy according to national guidelines in 49 patients (68.1%). Overtreated patients received on average more than two antibiotics. Main reasons for undertreatment were not administering antibiotics at all (n = 79, 22.3%) and failure to additionally administer Gentamicin (n = 217, 61.1%). For a detailed description, see *Appendix 8.C*.

Equivalently treated patients had comparable patient characteristics to adherently treated patients and thus an equal initial disease severity (NEWS of 6 (± 3.5) vs. 6 (± 3.8)). Only the number of patients in high triage categories was lower for equivalently treated patients (18.3% vs. 33.3%). Additionally, patients with underlying mild liver disease and malignancies were more frequently present in the equivalently treated group (*Table 8.2*).

Overtreated patients had characteristics that implied more critical illness than adherently treated patients. They were appointed to higher triage categories (52.4 % vs. 33.3%) and had worse vital signs. On average, overtreated patients had a higher NEWS of 8 (± 4.3) vs. 6 (± 3.8). Overtreatment more frequently occurred in patients with underlying chronic pulmonary disease (25.3% vs. 12.2%). Overtreated patients were more often diagnosed with mild- and moderate community-acquired pneumonia (*Table 8.2*).

Undertreated patients had characteristics that implied lower disease severity compared to adherently treated patients: they less frequently arrived by ambulance (16.6% vs. 30.1%), were less likely in high triage categories (14.7 % vs. 33.3%), and were less often directly admitted to the intensive care unit (2.0% vs. 11.6%). In addition, undertreated patients had more normal vital signs and on average a lower NEWS of 4 (± 3.1) vs. 6 (± 3.8). Especially, patients with underlying mild liver disease and chronic kidney disease were more often undertreated. Undertreated patients more often had a working diagnosis of cholangitis, pyelonephritis, and urosepsis (*Table 8.2*). We found that not administering recommended Gentamicin was more prevalent in patients with kidney disease (i.e. underlying chronic kidney disease and/or a suspected pyelonephritis/urosepsis. See *Appendix 8.C*.

Non-adherence and appropriate antibiotic coverage

Appropriate antibiotic coverage for the adherently treated was 89.1% (n = 352), for the equivalently treated 86.1% (n = 62), for the overtreated 94.3% (n = 82), and for the undertreated 58.0% (n = 206).

Equivalent treatment yielded comparable appropriate coverage compared to adherent treatment (OR[95%CI]: 0.86 [0.44 - 1.82], *Table 8.3*). Overtreatment did not result in higher appropriate coverage compared to adherent treatment (OR [95%CI]: 1.66 [0.77 - 4.16], *Table 8.3*). Undertreatment was associated with lower appropriate coverage compared to adherent treatment (OR[95%CI]: 0.18 [0.12 - 0.26], *Table 8.3*). After excluding patients that received no antibiotic therapy, appropriate coverage increased from 58.0% to 74.6%. However, undertreatment remained associated with lower appropriate coverage (OR [95%CI]: 0.27 [0.16 - 0.42]).

Table 8.2: Patient characteristics in adherently versus non-adherently treated patients

Characteristic	Adherence n = 395 (43.5)	Equivalent treatment n = 72 (7.9)	p-value	Overtreatment n = 87 (9.6)	p-value	Undertreatment n = 355 (39.1)	p-value
Sex, male	235 (59.5)	46 (63.9)	.48	57 (65.5)	.30	204 (57.5)	.57
Age, mean (SD)	61.0 (16.3)	58.9 (15.9)	.30	61.1 (14.4)	.99	61.6 (14.8)	.62
Arrival, by ambulance	119 (30.1)	19 (26.4)	.49	30 (34.5)	.43	59 (16.6)	< .001
Triage category, immediate/very urgent[a]	122 (33.3)	13 (18.3)	.01	44 (52.4)	.001	50 (14.7)	< .001
Direct intensive care unit admittance	46 (11.6)	5 (6.9)	.24	12 (13.8)	.58	7 (2.0)	< .001
Chills	156 (39.5)	38 (52.8)	.04	32 (36.8)	.64	169 (47.6)	.03
Vomiting	88 (22.3)	18 (25.0)	.61	23 (26.4)	.40	99 (27.9)	.08
Need for vasopressors	24 (6.1)	1 (1.4)	.10	10 (11.5)	.07	5 (1.4)	.001
Vital signs, mean (SD)							
Body temperature, °C[b]	38.4 (1.3)	38.4 (1.1)	.99	38.4 (1.0)	.59	38.2 (1.1)	.08
Heart rate, /min[c]	109 (24.4)	109 (23.3)	.98	117 (24.0)	.007	102 (20.6)	< .001
Respiratory rate, /min[d]	24 (8.6)	22 (7.2)	.11	27 (10.0)	.01	22 (7.1)	.003
Systolic blood pressure, mm Hg[e]	124 (28.5)	125 (35.2)	.77	120 (28.6)	.23	127 (24.6)	.08
Oxygen saturation, %[f]	95 (5.7)	96 (3.0)	.58	93 (9.9)	.07	96 (3.0)	.02
Any supplemental oxygen	195 (49.4)	35 (48.6)	.91	56 (64.4)	.01	99 (27.9)	< .001
Consciousness, not alert[g]	64 (18.8)	8 (12.1)	.20	19 (24.4)	.26	19 (6.8)	< .001
NEWS, mean (SD)[h]	6 (3.8)	6 (3.5)	.69	8 (4.3)	.001	4 (3.1)	< .001
Comorbidities							
Diabetes mellitus, uncomplicated	72 (18.2)	18 (25.0)	.18	14 (16.1)	.64	74 (20.8)	.37
Diabetes mellitus, end-organ damage	3 (0.8)	1 (1.4)	.59	3 (3.4)	.04	2 (0.6)	.74
Liver disease, mild	32 (8.1)	18 (25.0)	< .001	7 (8.0)	.99	69 (19.4)	< .001
Liver disease, moderate to severe	5 (1.3)	1 (1.4)	.93	0 (0.0)	.29	2 (0.6)	.32
Malignancy, leukemia, lymphoma, localized solid tumor	59 (14.9)	21 (29.2)	.003	6 (6.9)	.05	66 (18.6)	.18
Malignancy, metastatic solid tumor	48 (12.2)	15 (20.8)	.05	9 (10.3)	.64	50 (14.1)	.43

Table 8.2: Patient characteristics in adherently versus non-adherently treated patients (*continued*)

Characteristic	Adherence n = 395 (43.5)	Equivalent treatment n = 72 (7.9)	p-value	Overtreatment n = 87 (9.6)	p-value	Undertreatment n = 355 (39.1)	p-value
Chronic kidney disease	56 (14.2)	5 (6.9)	.09	13 (14.9)	.85	83 (23.4)	.001
Chronic pulmonary disease	48 (12.2)	6 (8.3)	.35	22 (25.3)	.002	45 (12.7)	.83
CCI, mean (SD)	4 (2.9)	5 (3.0)	.11	4 (2.9)	.84	5 (2.8)	.02
Origin, hospital acquired	205 (51.9)	46 (63.9)	.06	38 (43.7)	.17	199 (56.1)	.25
Ten most common working diagnoses							
Cholangitis	25 (6.3)	36 (50.0)	< .001	5 (5.7)	.84	115 (32.4)	< .001
Sepsis, urogenital	78 (19.7)	3 (4.2)	.001	7 (8.0)	.01	50 (14.1)	.04
Pyelonephritis	33 (8.4)	0 (0.0)	.01	1 (1.1)	.02	94 (26.5)	< .001
Sepsis, unknown	59 (14.9)	4 (5.6)	.03	10 (11.5)	.41	25 (7.0)	.001
Mild CA pneumonia (CURB 0-1)	12 (3.0)	6 (8.3)	.03	20 (23.0)	< .001	7 (2.0)	.35
Severe CA pneumonia (CURB 3-5)	18 (4.6)	1 (1.4)	.21	4 (4.6)	.99	14 (3.9)	.68
Febrile neutropenia	32 (8.1)	2 (2.8)	.11	1 (1.1)	.02	2 (0.6)	< .001
Meningitis, primary	24 (6.1)	0 (0.0)	.03	6 (6.9)	.77	2 (0.6)	< .001
Sepsis, abdominal	15 (3.8)	2 (2.8)	.67	0 (0.0)	.07	12 (3.4)	.76
Moderate CA pneumonia (CURB 2)	8 (2.0)	0 (0.0)	.22	16 (18.4)	< .001	2 (0.6)	.08

Data are presented as number (percentage) of patients unless otherwise indicated.
Missing data are not yet imputed.
Abbreviations: SD, standard deviation; NEWS, national early warning score; CCI, charlson comorbidity index; CURB65, confusion, blood urea nitrogen, respiratory rate, systolic blood pressure, age ≥ 65; CA, community-acquired.
[a]Data on triage category were missing for 44 (4.8%) patients.
[b]Data on body temperature were missing for 7 (0.8%) patients.
[c]Data on heart rate were missing for 18 (2.0%) patients.
[d]Data on respiratory rate were missing for 311 (34.2%) patients.
[e]Data on systolic blood pressure were missing for 14 (1.5%) patients.
[f]Data on oxygen saturation were missing for 39 (4.3%) patients.
[g]Data on consciousness were missing for 145 (16.0%) patients.
[h]NEWS imputed as normal.

Table 8.3: (Non-)adherence and appropriate antibiotic coverage

Type of (non-)adherence	Appropriate antibiotic coverage (%)	Odds ratio	95% CI
Adherence (n = 395)	352 (89.1)	1.0	(reference)
Non-adherence:			
Equivalent (n = 72)	62 (86.1)	0.86	0.44 - 1.82
Overtreatment (n = 87)	82 (94.3)	1.66	0.77 - 4.16
Undertreatment (n = 355)	206 (58.0)	0.18	0.12 - 0.26
Undertreatment, with no antibiotic therapy excluded (n = 276)	206 (74.6)	0.27	0.16 - 0.42

Non-adherence and 30-day mortality

Crude 30-day mortality for the adherently treated was 11.9% (n = 47), for the equivalently treated 13.9% (n = 10), for the overtreated 13.8% (n = 12), and for the undertreated 9.9% (n = 35). There was no association between the three types of non-adherence and 30-day mortality after both crude estimation and multivariable adjustment (OR [95%CI] ranging from: 0.65 [0.28 - 1.53] to 1.87 [0.79 - 4.41], *Table 8.4*). After excluding patients that received no antibiotic therapy, mortality rate for undertreatment decreased from 9.9% to 9.1%. However, undertreatment remained not associated with mortality (OR [95%CI]: 0.93 [0.52 - 1.89]).

Table 8.4: (Non-)adherence and 30-day mortality

Type of (non-)adherence	30-day mortality (%)	Crude odds ratio [95% CI]	Adjusted odds ratio[a] [95% CI]
Adherence (n = 395)	47 (11.9)	1.0 (reference)	1.0 (reference)
Non-adherence:			
Equivalent (n = 72)	10 (13.9)	1.19 [0.58 - 2.30]	1.87 [0.79 - 4.41]
Overtreatment (n = 87)	12 (13.8)	1.17 [0.59 - 2.17]	0.65 [0.28 - 1.53]
Undertreatment (n = 355)	35 (9.9)	0.82 [0.52 - 1.30]	1.16 [0.65 - 2.09]
Undertreatment, with no antibiotic therapy excluded (n = 276)	25 (9.1)	0.73 [0.44 - 1.20]	0.93 [0.52 - 1.89]

[a]Adjusted for: sex, age, arrival, triage category, direct intensive care unit admittance, chills, vomiting, vasopressors, body temperature, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, any supplemental oxygen, origin, consciousness, diabetes mellitus (uncomplicated), liver disease (mild), malignancy, chronic kidney disease, congestive heart failure, myocardial infarction, chronic pulmonary disease, perivascular disease, cerebrovascular accident, dementia, and connective tissue disease.

DISCUSSION

Our study aimed to evaluate non-adherence to antimicrobial guidelines for adult patients with BSI attending the ED. Non-adherence was high, and mostly the result of undertreatment. Non-adherence can be considered as equivalent, over-, or, undertreatment. As these are potentially distinctive groups with respect to severity of disease and potential adverse outcome, we analysed them separately. Previous studies did not stratify by type of non-adherence[211-213]. We found that, compared to adherently treated patients, overtreated patients were more severely ill, whilst undertreated patients were less severely ill. As a result, guideline adherence likely depends on clinical disease severity.

In the most severely ill patients, overtreatment may be a consequence of a physicians' intention to ensure appropriate antibiotic coverage. However, our study shows that this guideline deviation is not justifiable, because overtreatment was not associated with higher appropriate antibiotic coverage nor a survival benefit. Furthermore, overtreatment in general leads to risk of selection of antimicrobial resistance and adverse effects[208]. Therefore, adherence to the guidelines should be preferred to provide proper care, even when physicians encounter more severely ill patients. In accordance to previous studies, we found overtreatment was more frequent in patients with underlying chronic pulmonary disease and a suspected mild to moderate community-acquired pneumonia[225, 226]. Thus, for these patients with pulmonary disease, physicians should be extra alert to potential overtreatment.

In less severely ill patients, physicians can decide to give no or more narrow-spectrum antibiotic therapy than guidelines recommend. Undertreatment resulted in lower appropriate antibiotic coverage, also after excluding patients that received no antibiotic therapy. However, undertreatment was not associated with higher 30-day mortality in these patients with proven BSI. Although from our data there seems no survival disadvantage for these less severely ill undertreated patients, we have to emphasize that confounding by (low) severity of disease could mask potential survival disadvantages for the undertreated[227]. Therefore, physicians should always be cautious when they decide to undertreat and realize that appropriate antibiotic coverage is significantly lower compared to guideline-adherent therapy. We found undertreatment was more frequent in patients with underlying chronic kidney disease and a suspected pyelonephritis/urosepsis. From literature this can be explained by the intention to spare these patients from treatment with nephrotoxic antibiotics (i.e. failure to additionally administer Gentamicin)[228]. In our data, we found that not administering recommended Gentamicin was more prevalent in patients with kidney disease as well. Undertreatment was the

leading type of non-adherence, thus, antimicrobial guidelines often advise more extensive treatment than physicians in practice provide in less severely ill patients. For these patients, clinical judgement of low disease severity potentially overruled the guideline recommendations.

Non-adherence in our study (56.5%) was high compared to previously reported non-adherence rates (10 to 53%). However, previous studies are likely underestimating the true non-adherence rate as they excluded patients that received no antibiotic therapy[211-213]. Also, we chose to score adherence very strictly to give an unbiased interpretation of absolute guideline adherence. Strict scoring resulted in a few patients that were non-adherent, but equivalently treated with regard to antibiotic spectrum. Equivalently treated patients had comparable patient characteristics to adherently treated patients, indicating comparable illness. As expected, equivalent treatment yielded an equal rate of appropriate antibiotic coverage. Also, we found no difference in 30-day mortality.

Limitations

Our study has several limitations. First, we used retrospectively collected data making our study prone to bias. However, the quality of available data was assumed to be high as all data used was essential for daily clinical practice. For only 13 patients (1.3%) documentation was unclear on whether antibiotic therapy was administered in the ED or after discharge, therefore we scored them as no (and thus inappropriate) antibiotic coverage.

Also, we want to emphasize that we only considered empiric treatment in the ED, as this was our main study objective. Depending on disease course and culture results, antibiotic therapy could have been modified later on resulting in a different definitive antibiotic treatment. Aside from empiric antibiotic treatment in the ED, this may have altered survival as well.

CONCLUSIONS

In patients with BSI attending the ED, the majority of antibiotic therapy was non–adherent. Guideline adherence likely depends on clinical disease severity. Undertreatment was the leading type of non-adherence and was most common in less severely ill patients. Undertreatment was associated with lower appropriate antibiotic coverage, but not with higher mortality. However, it is important to realize that low severity of disease could mask survival disadvantages of undertreatment. Overtreatment was given to the most severely ill patients and did not result in higher appropriate antibiotic coverage

nor a survival benefit. Together with the risk of selection of antimicrobial resistance, overtreatment is not justifiable even in case of high disease severity.

Appendix 8.A: National Early Warning Score

We collected all vital signs of the National Early Warning Score (NEWS)[219]: body temperature, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, any supplemental oxygen, and consciousness (AVPU score: alert, voice, pain, unresponsive).

Each vital sign was graded 0-3. Scores for vital signs were added to obtain a total score. A NEWS over 7 triggers urgent clinical review[219].

Grading of vital signs	3	2	1	0	1	2	3
Body temperature, °C	< 35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	> 39.0	
Heart rate, beats/min	< 41		41 - 50	51 - 90	91 - 110	111 - 130	> 130
Respiratory rate, breaths/min	< 91	91 - 100	101 - 110	111 - 219			> 219
Systolic blood pressure, mmHg	< 9		9 - 11	12 - 20		21 - 24	> 25
Oxygen saturation, %	< 92	92 - 93	94 - 95	> 96			
Any supplemental oxygen		Yes		No			
Consciousness, AVPU				Alert			Not alert

Abbreviations: AVPU, alert, verbal, pain, unresponsive.

Appendix 8.B: Age-adjusted Charlson Comorbidity Index

We collected all comorbidities of the age-adjusted Charlson Comorbidity Index (CCI)[221]: diabetes mellitus (uncomplicated or end-organ damage), liver disease (mild or moderate to severe), malignancy (leukemia, lymphoma, localized solid tumor, or metastatic solid tumor), acquired immunodeficiency syndrome, chronic kidney disease, congestive heart failure, myocardial infarction, chronic pulmonary disease, peripheral vascular disease, cerebrovascular accident or transient ischemic attack, dementia, hemiplegia, connective tissue disease, and peptic ulcer disease.

Appendix 8.C: Detailed description of different types of guideline non-adherence

Type of non-adherence	N (%)
Undertreatment (n = 355)[c]	
Not administering recommended gentamicin[d]	217 (61.1)
Not administering any antibiotic therapy	79 (22.3)
Not administering recommended amoxicillin/clavulanic acid or metronidazole	26 (7.3)
Other reasons	33 (9.3)
Overtreatment (n = 87)[b]	
Not recommended administration of:	65 (74.7)
- One additional antibiotic agent	37
- Two additional antibiotic agents	19
- Three additional antibiotic agents	9
To broad-spectrum monotherapy	22 (25.3)
- Amoxicillin/clavulanic acid or levofloxacine or ciprofloxacine instead of amoxicillin	7
- Meropenem monotherapy (while not recommended)	6
- Amoxicillin/clavulanic acid or cefuroxim instead of flucloxacillin	4
- Piperacillin/tazobactam instead of cefuroxime and metronidazole	2
- Other reasons	3

Appendix 8.C: Detailed description of different types of guideline non-adherence (*continued*)

Type of non-adherence	N (%)
Equivalent treatment (n = 72)[a]	
Conform national guidelines:	49 (68.1)
- Amoxicillin/clavulanic acid and gentamicin is equivalent to cefuroxime and gentamicin (for cholangitis, unknown sepsis, and urosepsis)	40
- Amoxicillin/clavulanic acid is equivalent to ceftriaxone for primary peritonitis	5
- Amoxicillin/clavulanic acid is equivalent to cefuroxime and metronidazole	3
- Cephalosporin's are interchangeable	1
Not conform national guidelines:	23 (31.9)
- Amoxicillin is equivalent to cefuroxime for mild community-acquired pneumonia	6
- Amoxicillin/clavulanic acid is equivalent to cefuroxime and metronidazole	5
- Cephalosporin's are interchangeable	3
- Other reasons	9

Antibiotic therapy administered at the emergency department was with regard to antibiotic spectrum [a]equivalent, provided [b]higher coverage, or provided [c]lower coverage than guideline-recommended therapy.
[d] Not administering recommended gentamicin occurred more often in patients with underlying chronic kidney disease and/or a suspected pyelonephritis/urosepsis (38.5% vs. 17.9%, p < .001).

Appendix 8.D: Most frequently isolated bacteria

Isolated bacteria	N	Adherence n = 395	Undertreatment n = 355	Overtreatment n = 87	Equivalent treatment n = 72
Escherichia coli	311	117 (37.6)	151 (48.6)	18 (5.8)	25 (8.0)
Staphylococcus aureus	88	40 (45.5)	37 (42.0)	7 (8.0)	4 (4.5)
Streptococcus pneumoniae	85	48 (56.5)	14 (16.5)	21 (24.7)	2 (2.4)
Hemolytic streptococci (alpha, beta)	41	21 (51.2)	10 (24.4)	9 (22.0)	1 (2.4)
Other streptococcal species	41	20 (48.8)	12 (29.3)	6 (14.6)	3 (7.3)
Klebsiella species	82	31 (37.8)	40 (48.8)	2 (2.4)	9 (11.0)
Polymicrobial	60	29 (48.3)	23 (38.3)	3 (5.0)	5 (8.3)
Enterobacter species	35	13 (37.1)	18 (51.4)	1 (2.9)	3 (8.6)
Enterococcus species	17	3 (17.6)	10 (58.8)	1 (5.9)	3 (17.6)
Pseudomonas aeruginosa	23	11 (47.8)	7 (30.4)	4 (17.4)	1 (4.3)
Proteus mirabilis	17	10 (58.8)	4 (23.5)	3 (17.6)	0 (0.0)
Haemophilus influenzae	13	2 (15.4)	7 (53.8)	2 (15.4)	2 (15.4)
Citrobacter species	12	5 (41.7)	6 (50.0)	0 (0.0)	1 (8.3)
Staphylococcus epidermidis	11	7 (63.6)	0 (0.0)	4 (36.4)	0 (0.0)
Bacteroides fragilis	11	7 (63.6)	2 (18.2)	0 (0.0)	2 (18.2)
Neisseria meningitidis	4	3 (75.0)	0 (0.0)	1 (25.0)	0 (0.0)
Campylobacter species	4	3 (75.0)	1 (25.0)	0 (0.0)	0 (0.0)

chapter 9

The Association of Body Temperature with Antibiotic Therapy and Mortality in Patients Attending the Emergency Department with Suspected Infection

Romy Schuttevaer
Anniek Brink
Jelmer Alsma
Jurriaan E.M. de Steenwinkel
Annelies Verbon
Stephanie C.E. Schuit
Hester F. Lingsma

Eur J Emerg Med. 2021 Dec 1;28(6):440-447.

ABSTRACT

Background

Previous studies found that septic patients with normothermia have higher mortality than patients with fever. We hypothesize that antibiotic therapy is less frequently initiated if infectious patients present with normothermia to the emergency department (ED). We examined the association of body temperature with initiation of antibiotic therapy in patients attending the ED with suspected and proven infection. Additionally, the association of temperature with 30-day mortality was assessed.

Methods

We conducted a retrospective cohort study between 2012 and 2016 at a tertiary university hospital. Adult patients attending the ED with a blood culture taken (i.e. suspected infection) and a positive blood culture (i.e. proven bacteremia) were included. Tympanic temperature at arrival was categorized as hypothermia (< 36.1°C), normothermia (36.1 - 38.0°C), or hyperthermia (> 38.0°C). Primary outcome was initiation of antibiotic therapy. A secondary outcome was 30-day mortality. Multivariable logistic regression was used to control for covariates.

Results

Of 5,997 patients with a suspected infection 45.8% had normothermia, 44.6% hyperthermia, and 5.6% hypothermia. Patients with hyperthermia received more often antibiotic therapy (53.5%) compared to normothermic patients (27.6%, adjusted OR [95%CI]: 2.59 [2.27 - 2.95]). Patients with hyperthermia had lower mortality (4.7%) than those with normothermia (7.4%, adjusted OR [95%CI]: 0.50 [0.39 - 0.64]). Sensitivity analyses in patients with proven bacteremia (n = 934) showed similar results.

Conclusions

Normothermia in patients presenting with infection was associated with receiving less antibiotic therapy in the ED compared to presentations with hyperthermia. Moreover, normothermia was associated with a higher mortality risk than hyperthermia.

Keywords

Body temperature, Infection, Emergency department.

INTRODUCTION

Infections are potential life-threatening conditions with frequent presentations in an acute care setting such as the emergency department (ED)[205]. Body temperature, hereafter temperature, is an important vital sign, because fever is known to be a marker of infection[229]. Also, fever is a common reason to visit the ED[230]. Clinical decision support systems[231] such as the Manchester Triage System (MTS)[216] and National Early Warning Score (NEWS)[219] use temperature to assist in quickly identifying patients at high risk of deterioration in a general ED population. These decision support systems assign a higher mortality risk to hyperthermia (> 38.0°C) and hypothermia (< 36.1°C) compared to normothermia (36.1 - 38.0°C)[216, 219, 232]. However, according to a recent meta-analysis, septic patients with normothermia are at higher risk of dying than patients with hyperthermia[233]. It is unknown whether this is also generalizable to all patients attending the ED with a suspected infection.

We hypothesize that if patients with a suspected infection present to the ED with normothermia, antibiotic therapy may not be directly initiated. Instead these patients are first observed and antibiotic therapy is started when temperature becomes deviant or the patient deteriorates[234]. Such treatment delay would be unfavourable if normothermic patients with infection are at higher risk of dying than patients with hyperthermia and can be in contrast to the recommendations of the surviving sepsis campaign regarding timing of antibiotic treatment[235].

The primary aim of this study was to examine the association of temperature with initiation of antibiotic therapy in patients attending the ED with a suspected infection (i.e. with a blood culture taken) and in patients with proven bacteraemia (i.e. with a positive blood culture). A secondary endpoint was 30-day mortality.

METHODS

Study design and setting

We conducted a retrospective cohort study at the Erasmus University Medical Center Rotterdam (Erasmus MC), which is the largest tertiary referral centre in the Netherlands. The ED is an open access department with approximately 33,000 annual visits. Our database consists of automatically derived data from all patients admitted to the ED between July 1st 2012 and December 31st 2016. The Medical Ethics Committee of the Erasmus MC reviewed the study and concluded that our study did not fall under the scope of the Medical Research Involving Human Subjects Act and therefore no informed consent needed to be obtained. Our study is registered under MEC-2018-1744 and MEC-205-106.

Selection of participants

Patients were eligible for inclusion if they were at least 18 years of age and had a blood culture taken in the ED (i.e. a suspected infection). Blood cultures were obtained prior to administering antibiotic therapy (in line with the surviving sepsis campaign[235]) in patients with presentations suggestive of infection (i.e. based on history, vital signs, inflammation parameters[236]) requiring intravenous antibiotics, in patients that were prone for a severe course of infection (e.g. patients with immunodeficiency), or in severely ill patients with undifferentiated presentations.

Subsequently, we selected a subgroup of patients in which blood cultures were positive (i.e. proven bacteraemia). Only the first ED visit of both the 'blood culture taken'-group and the 'positive blood culture'-group was included to prevent domination of results by a small group of individuals that frequently visited the ED.

Data collection

Data were derived from the ED and were combined with a database from Medical Microbiology, containing all collected blood cultures[220]. The ED database was collected automatically with text mining from patient charts and consists of first recorded tympanic temperature, antibiotic therapy initiated in the ED, and mortality which was updated from municipal death registration records. Collected patients characteristics were: demographics (age, sex), arrival mode (by ambulance), triage category (according to the MTS)[216], other first recorded vital signs (i.e. systolic blood pressure, respiratory rate, peripheral oxygen saturation, consciousness[237]), whether there was need for any supplemental oxygen, and hospital admittance. Additionally we collected the inflammation parameters: C-reactive protein (CRP) and leukocyte count. We had no data on the exact time to antibiotic administration in the ED. However, if antibiotics were prescribed for patients that were hospitalised the first dose was always given in the ED.

The Medical Microbiology database contains data about the type of pathogen for all positive blood cultures in the ED. A positive blood culture was defined as presence of a known pathogen (e.g. E. coli) in one blood culture or a less pathogenic bacteria (e.g. S. epidermidis)[238] in at least two blood cultures collected on separate occasions within two days from ED admission[215, 238]. For all patients with a positive blood culture, we manually reviewed patient charts and additionally collected: disposition to the intensive care unit directly from the ED, chills[217], vomiting[217], need for vasopressors, and the age adjusted Charlson Comorbidity Index (CCI)[221].

Data processing

We handled missing data with multiple imputations using the chained equations method with five datasets[239]. To provide constant temperature groups, temperature was single imputed. Temperature at arrival was categorized as hypothermia (< 36.1°C), normothermia (36.1 - 38.0°C), or hyperthermia (> 38.0°C). This categorization corresponds to previous studies and the cut-off values of NEWS[219, 233]. Additionally, we calculated the NEWS, because this is proposed as an accurate early warning system in patients with suspected infection in the ED[232]. The initiated intravenous antibiotic therapy in the ED was recoded while accounting for potential typing errors, abbreviations, and brand names (e.g. Cefuroxime was scored if either 'cefur' or 'zinacef' was documented).

Data analysis

We described arrival temperature (i.e. normothermia, hyperthermia, hypothermia, or missing) and proportions of 30-day mortality for our total ED population, for patients with and without a blood culture taken, and for patients with- and without a positive blood culture. All subsequent analyses were performed in patients with a blood culture taken and for sensitivity analysis repeated in the subgroup of patients with a positive blood culture.

We examined presenting patient characteristics among patients with normothermia, hyperthermia and hypothermia. Based on distribution, data were compared using unpaired t-tests, chi-squared tests, Mann-Whitney U test, or Fisher's exact tests.

We investigated the association between temperature (i.e. hyperthermia compared to normothermia and hypothermia compared to normothermia) and [1] initiation of antibiotic therapy in the ED, [2] 30-day mortality, [3] hospital admittance, and [4] blood culture positivity. Thus, our reference category was normothermia. For the association between temperature and [2] 30-day mortality we analysed temperature as a continuous variable as well, and investigated whether there was a U-shaped association (as proposed by MTS and NEWS[219, 231, 232]) or an inverse association (as proposed by a recent meta-analysis[233]) with use of restricted cubic splines (with 3 knots). For all associations, we used multivariable logistic regression to control for covariates (i.e. sex, age, triage category, vital signs, inflammation parameters, for detailed information see *Appendix 9.A*).

Results were presented as odds ratios (OR) with 95% confidence intervals (CI). All hypothesis tests were 2-sided, with a significance level of $p < .05$. Statistical analyses were performed using R version 4.0.1.

RESULTS

Of 65 986 unique adult patients that visited our ED between July 1st 2012 and December 31st 2016 we included 5,997 patients with a blood culture taken (i.e. suspected infection) in the ED (*Figure 9.1*). For the sensitivity analysis, 934 patients with a positive blood culture (i.e. proven bacteraemia) were selected (*Figure 9.1*).

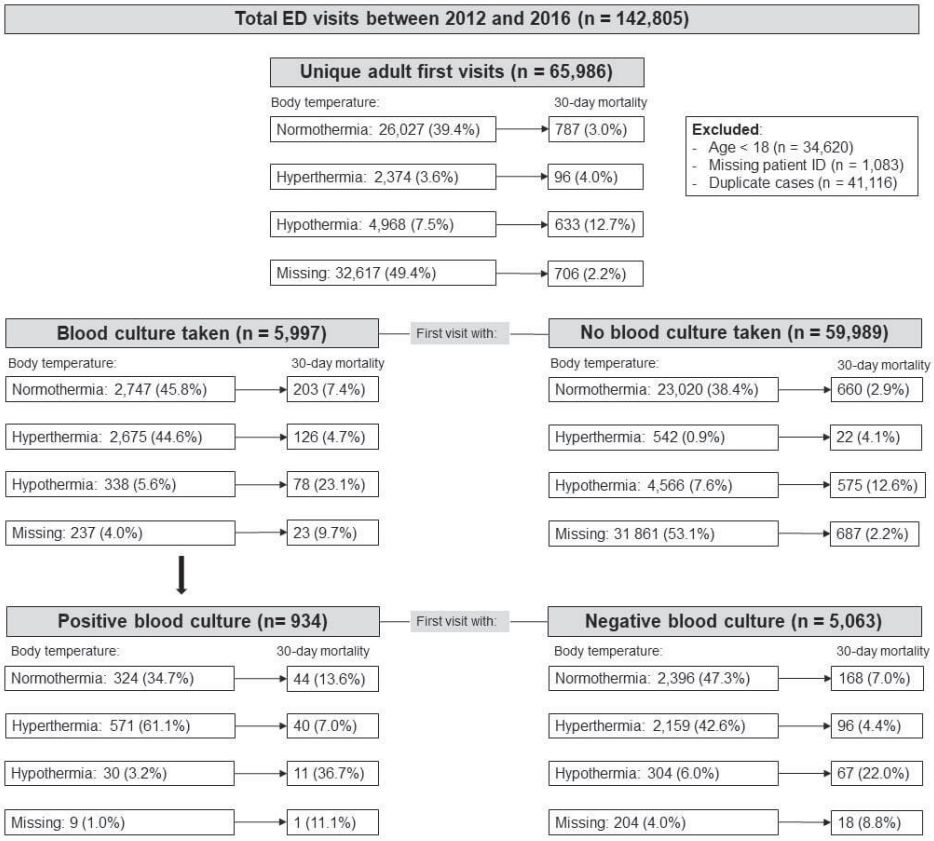


Figure 9.1: Flowchart of study selection and mortality (30-day).
Abbreviations: ED, emergency department.
For all selection steps first visits were included and therefore some patients in the 'blood culture taken'-group and 'positive blood culture'-group are included at a later point in time (e.g. a few patients are included in the 'blood culture taken'-group, but had a positive blood culture in a subsequent ED visit and are therefore in the 'positive blood culture'-group later in time).

Patient characteristics

In all 5,997 patients with suspected infection: 45.8% had normothermia, 44.6% hyperthermia, and 5.6% hypothermia. Temperature was missing in 4.0% (*Figure 9.1*). Normothermic patients were less frequently categorized as urgent (11.4%) than patients

with hyperthermia (17.5%) or hypothermia (46.2%). The average NEWS was higher for patients with hyperthermia (6 ± 3.4) compared to patients with normothermia (4 ± 3.3), but comparable if temperature and heart rate were not incorporated (3 ± 2.9 for hyperthermia versus 3 ± 2.9 for normothermia). Patients with hypothermia had a higher NEWS than patients with normothermia, also without incorporating temperature and heart rate (5 ± 3.7 for hypothermia versus 3 ± 2.9 for normothermia). CRP and leukocyte count did not differ between normothermic patients and patients with hyperthermia. Hypothermic patients had more deviating inflammation parameters (*Table 9.1*).

Table 9.1: Patient characteristics in normothermia, hyperthermia and hypothermia for patients with a blood culture taken in the ED - i.e. suspected infection (n = 5,997)

Characteristic	Missing	Normothermia n = 2,747 (45.8)	Hyperthermia n = 2,675 (44.6)	p-value hyper versus normothermia	Hypothermia n = 338 (5.6)	p-value hypo versus normothermia
Sex, male	0	1,528 (55.6)	1,509 (56.4)	0.56	225 (66.6)	< 0.001
Age, mean (SD)	0	56 (17.2)	55 (17.8)	0.11	63 (15.6)	< 0.001
Arrival, by ambulance	0	324 (11.8)	396 (14.8)	0.001	135 (39.9)	< 0.001
Triage by MTS, acute/highly urgent	271 (4.5)	302 (11.4)	450 (17.5)	< 0.001	145 (46.2)	< 0.001
Vital signs, mean (SD):						
Temperature, °C	237 (4.0)	37.2 (0.5)	38.8 (0.6)	< 0.001	35.1 (1.3)	< 0.001
Heart rate, /min	294 (4.9)	95 (20.0)	107 (19.2)	< 0.001	89 (26.7)	< 0.001
Respiratory rate, /min	2435 (40.6)	21 (7.2)	22 (7.7)	< 0.001	21 (8.2)	0.15
Systolic blood pressure, mm Hg	388 (6.5)	131 (23.4)	134 (24.0)	< 0.001	127 (34.0)	0.08
Oxygen saturation, %	424 (7.1)	96 (3.7)	96 (3.3)	0.002	95 (5.1)	0.04
Any supplemental oxygen	0	764 (27.8)	952 (35.6)	< 0.001	177 (52.4)	< 0.001
Consciousness, not alert	1,251 (20.9)	286 (13.0)	338 (15.5)	0.02	111 (39.8)	< 0.001
NEWS, mean (SD)*	0	4 (3.3)	6 (3.4)	< 0.001	7 (4.4)	< 0.001
NEWS without temperature, mean (SD)*	0	4 (3.3)	5 (3.3)	< 0.001	6 (4.1)	< 0.001
NEWS without temperature and heart rate, mean (SD)*	0	3 (2.9)	3 (2.9)	< 0.001	5 (3.7)	< 0.001
Inflammation parameters, median (IQR):						
CRP, mg/L	998 (16.6)	58 (120.4)	58 (91.7)	0.93	19 (88.9)	< 0.001
Leukocyte count, × 10^3/μL	1,024 (17.1)	7.7 (11.5)	5.0 (9.6)	0.90	13.2 (15.1)	0.07

Normothermia (36.1-38.0°C) is compared to hyperthermia (> 38.0°C) and hypothermia (< 36.1°C).
Data are presented as number (percentage) of patients unless otherwise indicated.
Data in this table are unimputed.
Abbreviations: MTS, Manchester triage system; NEWS, national early warning score; CRP, C-reactive protein.
*NEWS imputed as normal.

In the sensitivity analysis limited to 934 patients with a positive blood culture: 34.7% had normothermia, 61.1% hyperthermia, and 3.2% hypothermia. Temperature was missing in 1.0% (*Figure 9.1*). Most patient characteristics were comparably distributed to patients with a blood culture taken (*Appendix 9.B*). However, compared to patients with hyperthermia, normothermic patients needed vasopressors more frequently (2.5% versus 5.2%). CRP and leukocyte count were higher for normothermic patients (respectively median 207 mg/L and 7.8 × 10^3/μL) than for patients with hyperthermia (respectively median 85 mg/L and 3.7 × 10^3/μL**)**. Thus, normothermic patients with a positive blood culture appeared more seriously ill than patients with hyperthermia. Comorbidity, expressed as the CCI, was equal for each temperature group (*Appendix 9.B*).

Temperature and initiation of antibiotic therapy

In all 5,997 patients with suspected infection antibiotic therapy was more often initiated in the ED if patients had hyperthermia or hypothermia (respectively 53.5%, 47.6%) compared to normothermic patients (27.6%, crude OR [95%CI] for hyperthermia: 2.96 [2.65 - 3.31] and for hypothermia 2.21 [1.77 - 2.75], *Table 9.2*). These associations were independent of covariates (adjusted OR [95%CI] for hyperthermia: 2.59 [2.27 - 2.95] and for hypothermia 1.42 [1.08 - 1.87], *Table 9.2*).

In the sensitivity analysis limited to 934 patients with a positive blood culture, antibiotic therapy was also more often administered in patients with hyperthermia or hypothermia (respectively 86.9% and 93.3%) compared to patients with normothermia (72.2%, crude OR [95%CI] for hyperthermia: 2.64 [1.88 - 3.73] and for hypothermia 5.54 [1.29 - 23.84], *Table 9.2*). After adjustment, these associations subsisted (adjusted OR [95%CI] for hyperthermia: 2.40 [1.59 - 3.61] and for hypothermia 5.91 [1.00 - 32.86], *Table 9.2*).

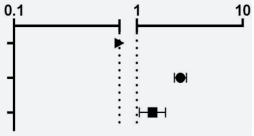
Temperature and mortality (30-day)

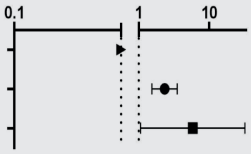
In all 5,997 patients with suspected infection, mortality rate was higher for normothermic patients (7.4%) than for patients with hyperthermia (4.7%, *Figure 9.1*). There was an inverse association between temperature and mortality rather than a U-shaped association (*Figure 9.2*). An increasing temperature was associated with lower 30-day mortality, both crude (OR per degree increase [95%CI]: 0.66 [0.62 - 0.71]) and after adjustment for covariates (OR per degree increase [95%CI]: 0.74 [0.69 - 0.81], *Table 9.3*).

In the sensitivity analysis limited to 934 patients with a positive blood culture (n = 934), normothermic patients had a higher mortality rate (13.6%) compared to those with hyperthermia (7.0%). An increasing temperature was associated with lower mortality both crude (OR per degree increase [95%CI]: 0.62 [0.52 - 0.74]) and after adjustment for covariates (OR per degree increase [95%CI]: 0.72 [0.58 - 0.89], *Table 9.3*).

Table 9.2: Temperature and initiation of antibiotic therapy

Temperature	N	Antibiotics (%)	Crude OR [95% CI]	Adjusted OR* [95% CI]
Blood culture taken†	5,997			
Normothermia (36.1-38.0 ° C)	2,747	758 (27.6)	1.0 (reference)	1.0 (reference)
Hyperthermia (> 38.0 ° C)	2,675	1,430 (53.5)	2.96 [2.65 - 3.31]	2.59 [2.27 - 2.95]
Hypothermia (< 36.1 ° C)	338	161 (47.6)	2.21 [1.77 - 2.75]	1.42 [1.08 - 1.87]
Positive blood culture‡	934			
Normothermia (36.1-38.0 ° C)	324	234 (72.2)	1.0 (reference)	1.0 (reference)
Hyperthermia (> 38.0 ° C)	571	496 (86.9)	2.64 [1.88 - 3.73]	2.40 [1.59 - 3.61]
Hypothermia (< 36.1 ° C)	30	28 (93.3)	5.54 [1.29 - 23.84]	5.91 [1.00 - 32.86]





Hyperthermia (> 38.0°C) and hypothermia (< 36.1°C) are compared to normothermia (36.1-38.0°C).
Data on number and percentages are unimputed, odds ratios are obtained from imputed data.
*Adjusted for: sex, age, arrival, triage category, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, any supplemental oxygen, consciousness, CRP, and leukocyte count.
†Data on temperature were missing for 237 (4.0%) patients in which antibiotic were administered in 31 (13.1%). Excluding these patients did not affect the results.
‡Data on temperature were missing for 9 (1.0%) patients in which antibiotics were administered in 6 (66.7%). Excluding these patients did not affect the results.
Data on number and percentages are unimputed, odds ratios are obtained from imputed data.
*Adjusted for: sex, age, arrival, triage category, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, any supplemental oxygen, consciousness, CRP, and leukocyte count.
†Data on temperature were missing for 237 (4.0%) patients in which antibiotic were administered in 31 (13.1%). Excluding these patients did not affect the results.
‡Data on temperature were missing for 9 (1.0%) patients in which antibiotics were administered in 6 (66.7%). Excluding these patients did not affect the results.

For both the total of ED visits (n = 65,986) and for patients without a blood culture taken (n = 59,989) mortality rates were higher for patients with hyperthermia (4.0-4.1%) than for patients with normothermia (2.9-3.0%), which corresponds to scoring by MTS and NEWS (*Figure 9.1*).

Temperature and hospital admittance

In all patients with suspected infection, hyperthermia resulted more often in hospital admittance (83.8%) than normothermia (73.5%, *Appendix 9.C*).

In the sensitivity analysis limited to patients with a positive blood culture, there were no statistically significant differences in hospital admittance between temperature groups (*Appendix 9.C*).

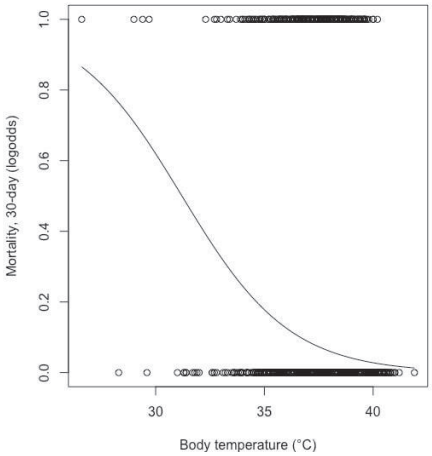
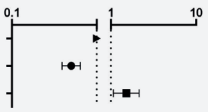
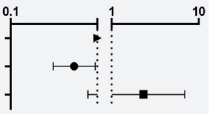


Figure 9.2: Unadjusted probability of mortality (30-day). There was an inverse association between temperature and mortality rather than a U-shaped association. Modelling temperature as a restricted cubic spline (3 knots) did not improve the model. Lower temperature was associated with a higher risk of mortality

Table 9.3: Temperature and mortality (30-day)

Temperature	N	30-day mortality (%)	Crude OR [95% CI]	Adjusted OR* [95% CI]
Blood culture taken	5,997			
Normothermia† (36.1-38.0 ° C)	2,747	203 (7.4)	1.0 (reference)	1.0 (reference)
Hyperthermia (> 38.0 ° C)	2,675	126 (4.7)	0.57 [0.46 - 0.72]	0.50 [0.39 - 0.64]
Hypothermia (< 36.1 ° C)	338	78 (23.1)	3.56 [2.69 - 4.71]	1.52 [1.07 - 2.15]
Temperature, °C (continuous)‡	5,760	407 (7.1)	0.66 [0.62 - 0.71]	0.74 [0.69 - 0.81]
Positive blood culture	934			
Normothermia§ (36.1-38.0 ° C)	324	44 (13.6)	1.0 (reference)	1.0 (reference)
Hyperthermia (> 38.0 ° C)	571	40 (7.0)	0.47 [0.30 - 0.74]	0.54 [0.31 - 0.94]
Hypothermia (< 36.1 ° C)	30	11 (36.7)	3.64 [1.62 - 8.18]	2.35 [0.79 - 7.00]
Temperature, °C (continuous)‖	925	95 (10.3)	0.62 [0.52 - 0.74]	0.72 [0.58 - 0.89]





Hyperthermia (> 38.0°C) and hypothermia (< 36.1°C) are compared to normothermia (36.1-38.0°C).
Data on number and percentages are unimputed, odds ratios are obtained from imputed data.
*Adjusted for: sex, age, arrival, triage category, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, any supplemental oxygen, consciousness, CRP, leukocyte count, and antibiotic therapy.
Modelling temperature as a restricted cubic spline (3 knots) did not improve the crude and adjusted models for both patients with a blood culture taken‡ and patients with a positive blood culture‖.
‡Data on temperature were missing for 237 (4.0%) patients in which 30-day mortality was 23 (9.7%). Excluding these patients did not affect the results.
‖Data on temperature were missing for 9 (1.0%) patients in which 30-day mortality was 1 (11.1%). Excluding these patients did not affect the results.

Temperature and blood culture positivity

Blood cultures were positive in 11.9% of patients with normothermia and in 20.9% of patients with hyperthermia. An increasing temperature was associated with blood culture positivity both crude (OR [95%CI]: 1.46 [1.36 - 1.40], *Figure 9.3*) and after adjustment for covariates (OR [95%CI]: 1.48 [1.37 - 1.60], *Appendix 9.D*).

Negative blood cultures

In patients with negative blood cultures (n = 5,063), comparable rates of antibiotic administration, 30-day mortality, and hospital-admittance for normothermia, hyperthermia, and hypothermia were found as for patients with a (positive) blood culture taken (*Appendix 9.E*).

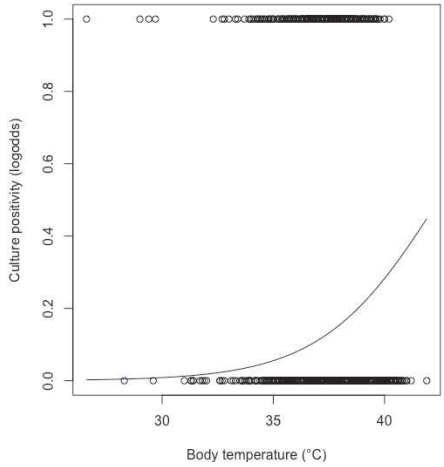


Figure 9.3: culture-positivity for temperature in patients with a blood culture taken. Among patients with a blood culture taken, the probability of a positive blood culture increased with an increasing temperature.

DISCUSSION

In this study, the association of temperature with initiation of antibiotic therapy and additionally 30-day mortality was addressed in patients attending the ED with a suspected infection (i.e. blood culture taken). Normothermic infections were common and antibiotic therapy was significantly less frequently initiated if patients presented with normothermia, compared to presentations with hyperthermia or hypothermia. However, normothermia was associated with higher 30-day mortality than hyperthermia, as has been described by others[55, 233, 240]. Patients with hypothermia had highest mortality risk. Moreover, in patients with proven bacteraemia (i.e. positive blood cultures), normothermia implied higher disease severity and yet these patients received less antibiotic therapy.

Higher mortality among normothermic patients with infection has multiple explanations. One explanation is that normothermic patients are potentially incorrectly assessed as lower acuity because fever is lacking. Which can result in delay in diagnosis and initiation of antibiotic therapy if these patients are first observed (i.e. "watchful waiting") and antibiotic therapy is started only when temperature becomes deviant or the patient deteriorates. These delays are understandable because hyperthermia is considered as a marker of infection[241]. However, hyperthermia is a poor predictor of mortality among infectious patients as an increasing temperature was associated with lower 30-day mortality (i.e. an inverse association). Clinical decision support systems such as NEWS[219] and MTS[216] assign a higher mortality risk to hyperthermia than to normothermia. This seems appropriate for a general ED population, but not for patients with infection.

To examine to what extent higher mortality among normothermic patients with infection is attributable to the lack of antibiotic treatment can only fairly be studied in research with prospective designs. Retrospective data does not allow to study the effects of antibiotic therapy on mortality, because there is a high risk of bias due to confounding by indication (i.e. patients that are already at high risk of dying are more likely to receive antibiotic therapy than lower acuity patients)[227, 242]. Nonetheless, it is likely to assume that mortality among patients with infection could at least for some extent have been reduced, would these patients have had early initiation of antibiotic therapy in the ED[235].

Aside from inadequate recognition of disease severity and subsequent lack of antibiotic treatment, there are other potential explanations for higher mortality among normothermic patients with infection. Normothermic patients may have an impaired febrile response to infection due to older age, comorbidity, use of anti-pyretic drugs, or because of more critical acute illness[241]. As a result, normothermic patients might represent a patient group that is older and has higher disease severity than patients with hyperthermia. There were no differences in age or comorbidity, however, among normothermic patients with bacteraemia, CRP and leukocyte levels were higher compared to patients with hyperthermia. Also, these normothermic patients needed vasopressors more frequently in the ED. Consequently, physicians should be aware that in patients with infection normothermia is not a sign of minor disease but may even imply a more serious course of illness. If the only drawback of initiation of antibiotic therapy is the absence of fever, physicians should more often reconsider starting.

Limitations

This study has some limitations. Retrospectively collected data was used which makes it prone to bias[227]. However, the quality of available data was high as all data used was

essential for daily clinical practice. To preserve generalizability, patients with a blood culture taken in the ED were selected, because this is the point in time in which antibiotic therapy should be initiated if indicated according to the surviving sepsis campaign[235]. However, patients with a blood culture taken may also resemble non-infectious pathology. Therefore, analyses were repeated in the subgroup of patients with proven bacteraemia, which is a group that retrospectively had a true bacterial infection. Also, blood cultures were taken in case of a certain suspicion of infection (e.g. sepsis), which potentially ruled out patients with localized infections and therefore our results are not generalizable to this group. Another limitation of this study is that only admission temperature was accessible and there were no data on use of antipyretic drugs prior to the ED visit. Additionally, in our research setting (the ED), it was not possible to obtain all SOFA criteria and therefore our population could not formerly be defined as being septic or not according to the sepsis-3 definitions. We were unable to study the association between temperature and delays in antibiotic administration, because there was no data on the exact time to antibiotic administration in the ED. However, if antibiotics are prescribed for patients that are hospitalised, the first dose is always given in the ED independent from ED length of stay. Moreover, it would be interesting to study the effects of antibiotic therapy on mortality among normothermic patients with infection. However, we were unable to examine this with our retrospective study design because of the high risk of bias due to confounding by indication (i.e. patients that are already at high risk of dying are more likely to receive antibiotic therapy than lower acuity patients)[227, 242]. Additionally, the possibility of missing cases receiving antibiotic therapy in patients with suspected infection cannot be excluded due to data collection with automatic text mining. However, for patients with proven bacteraemia patient charts were manually reviewed and comparing both ways of data collection did not affect the results.

CONCLUSION

In this retrospective cohort study in patients presenting with infection, normothermia was associated with receiving less antibiotic therapy in the ED compared to presentations with hyperthermia. Moreover, normothermia was associated with a higher mortality risk than hyperthermia. Physicians should be aware that normothermia does not exclude infection and may even imply a more serious course of illness.

Appendix 9.A: Data analysis: covariate adjustment

Association between temperature and:	Covariates that were multivariably adjusted for:
Antibiotic therapy	Sex, age, arrival, triage category, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, any supplemental oxygen, consciousness, CRP, and leukocyte count
30-day mortality, hospital admittance, blood culture positivity	Sex, age, arrival, triage category, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, any supplemental oxygen, consciousness, CRP, leukocyte count, and antibiotic therapy

Appendix 9.B: Patient characteristics in normothermia, hyperthermia and hypothermia for patients with a positive blood culture - i.e. proven bacteremia (n = 934)

Characteristic	Missing	Normothermia n = 324 (34.7)	Hyperthermia n = 571 (61.1)	p-value hyper vs. normothermia	Hypothermia n = 30 (3.2)	p-value hypo vs. normothermia
Sex, male	0	198 (61.1)	351 (61.5)	0.92	18 (60.0)	0.91
Age, mean (SD)	0	61 (14.5)	60 (16.0)	0.18	67 (13.6)	0.04
Arrival, by ambulance	0	49 (15.1)	121 (21.2)	0.03	19 (63.3)	< 0.001
Triage by MTS, acute/highly urgent	48 (5.1)	47 (15.3)	149 (27.0)	< 0.001	16 (55.2)	< 0.001
Direct intensive care unit admittance	0	22 (6.8)	39 (6.8)	0.98	2 (6.7)	0.98
Chills	0	133 (41.0)	277 (48.5)	0.03	6 (20.0)	0.02
Vomiting	0	76 (23.5)	145 (25.4)	0.52	8 (26.7)	0.69
Need for vasopressors	0	17 (5.2)	14 (2.5)	0.03	5 (16.7)	0.01
Vital signs, mean (SD)						
Temperature, °C	9 (1.0)	37.3 (0.5)	39.0 (0.6)	< 0.001	35.1 (1.8)	< 0.001
Heart rate, /min	23 (2.5)	100 (21.1)	110 (21.4)	< 0.001	97 (34.4)	0.63
Respiratory rate, /min	328 (35.1)	22 (7.9)	23 (8.6)	0.02	23 (8.7)	0.32
Systolic blood pressure, mm Hg	20 (2.1)	120 (26.9)	129 (26.8)	< 0.001	108 (26.7)	0.02
Oxygen saturation, %	39 (4.2)	96 (5.8)	96 (3.5)	0.90	94 (4.6)	0.19
Any supplemental oxygen	0	108 (32.9)	240 (41.7)	0.009	16 (53.3)	0.03
Consciousness, not alert	172 (18.4)	24 (9.5)	62 (13.1)	0.15	9 (30.0)	< 0.001
NEWS, mean (SD)*	0	4 (3.6)	6 (3.5)	< 0.001	8 (3.9)	< 0.001
NEWS without temperature, mean (SD)*	0	4 (3.6)	5 (3.4)	0.003	6 (3.6)	0.001
NEWS without temperature and heart rate, mean (SD)*	0	3 (3.2)	3 (3.1)	0.19	5 (2.8)	< 0.001
CCI, mean (SD)†	0	4 (2.9)	4 (2.9)	0.75	5 (2.8)	0.19
Suspected source of infection:						

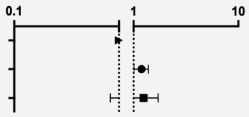
Appendix 9.B: Patient characteristics in normothermia, hyperthermia and hypothermia for patients with a positive blood culture - i.e. proven bacteremia (n = 934) (*continued*)

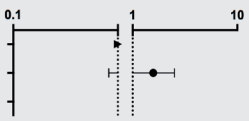
Characteristic	Missing	Normothermia n = 324 (34.7)	Hyperthermia n = 571 (61.1)	p-value hyper vs. normothermia	Hypothermia n = 30 (3.2)	p-value hypo vs. normothermia
Abdominal	0	87 (26.9)	142 (24.9)	.51	9 (30.0)	.71
Urogenital	0	74 (22.8)	155 (27.1)	.16	6 (20.0)	.72
Respiratory	0	40 (12.3)	75 (13.1)	.74	4 (13.3)	.88
Inflammation parameters, median (IQR):						
CRP, mg/L	137 (14.7)	207 (157.3)	85 (113.4)	< 0.001	272 (131)	< 0.001
Leukocyte count, × 10^3/µL	145 (15.5)	7.8 (14.5)	3.7 (10.5)	0.006	13.6 (13.8)	.15
Isolated bacteria						
Escherichia coli	0	99 (30.6)	198 (34.7)	.21	9 (30.0)	.95
Staphylococcus aureus	0	35 (10.8)	50 (8.8)	.32	6 (20.0)	.13
Klebsiella pneumoniae	0	18 (5.6)	57 (10.0)	.02	2 (6.7)	.80

Normothermia (36.1-38.0°C) is compared to hyperthermia (> 38.0°C) and hypothermia (< 36.1°C).
Data are presented as number (percentage) of patients unless otherwise indicated.
Data in this table are unimputed.
Abbreviations: MTS, Manchester triage system; NEWS, national early warning score; CCI, charlson comorbidity index; CRP, C-reactive protein.
*NEWS imputed as normal.
†Individual comorbidities of the CCI were not differing between temperature groups.

Appendix 9.C: Temperature and hospital admittance

Temperature	N	Admittance (%)	Crude OR [95% CI]	Adjusted OR* [95% CI]
Blood culture taken†	5,997			
Normothermia (36.1-38.0 ° C)	2,747	2,019 (73.5)	1.0 (reference)	1.0 (reference)
Hyperthermia (> 38.0 ° C)	2,675	2,241 (83.8)	1.78 [1.57-2.03]	1.18 [1.01-1.38]
Hypothermia (< 36.1 ° C)	338	292 (86.4)	2.20 [1.62-2.99]	1.20 [0.84-1.72]
Positive blood culture‡	934			
Normothermia (36.1-38.0 ° C)	324	289 (89.2)	1.0 (reference)	1.0 (reference)
Hyperthermia (> 38.0 ° C)	571	529 (92.6)	1.58 [0.99-2.50]	1.40 [0.82-2.51]
Hypothermia (< 36.1 ° C)	30	30 (100)	∞	∞

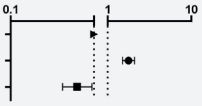




Hyperthermia (> 38.0°C) and hypothermia (< 36.1°C) are compared to normothermia (36.1-38.0°C).
Temperature was single imputed to provide constant temperature groups, other data are multiply imputed.
Abbreviations: OR, odds ratio; CI, confidence interval.
*Adjusted for: sex, age, arrival, triage category, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, any supplemental oxygen, consciousness, CRP, leukocyte count, and antibiotic therapy.
†Data on temperature were missing for 237 (4.0%) patients.
‡Data on temperature were missing for 9 (1.0%) patients.

Appendix 9.D: Temperature and blood culture positivity

Temperature	N	Positive blood culture (%)	Crude OR [95% CI]	Adjusted OR* [95% CI]
Blood culture taken†	5,997			
Normothermia (36.1-38.0 ° C)	2,720	324 (11.9)	1.0 (reference)	1.0 (reference)
Hyperthermia (> 38.0 ° C)	2,730	571 (20.9)	1.92 [1.66 - 2.22]	1.80 [1.53 - 2.11]
Hypothermia (< 36.1 ° C)	334	30 (9.0)	0.82 [0.57 - 1.19]	0.62 [0.41 - 0.93]
Temperature, °C (continuous)	5,784	925 (16.0)	1.46 [1.36 - 1.40]	1.48 [1.37 - 1.60]



Hyperthermia (> 38.0°C) and hypothermia (< 36.1°C) are compared to normothermia (36.1-38.0°C).
Data on number and percentages are unimputed, odds ratios are obtained from imputed data.
*Adjusted for: sex, age, arrival, triage category, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, any supplemental oxygen, consciousness, CRP, leukocyte count, and antibiotic therapy.
†Data on temperature were missing for 237 (4.0%) patients.

Appendix 9.E: Negative blood cultures

Temperature	N	Antibiotics (%)	30-day mortality (%)	Admittance (%)
Negative blood cultures	5,063			
Normothermia (36.1-38.0 ° C)	2,396	607 (25.3)	168 (7.0)	1,718 (71.7)
Hyperthermia (> 38.0 ° C)	2,159	1,085 (50.3)	96 (4.4)	1,764 (81.7)
Hypothermia (< 36.1 ° C)	304	136 (44.7)	67 (22.0)	258 (84.9)


chapter
10

Appropriate Empirical Antibiotic Therapy and Mortality: Conflicting Data Explained by Residual Confounding

Romy Schuttevaer
Jelmer Alsma
Anniek Brink
Willian van Dijk
Jurriaan E.M. de Steenwinkel
Hester F. Lingsma
Damian C. Melles
Stephanie C.E. Schuit

PLoS One. 2019 Nov 19;14(11):e0225478.

ABSTRACT

Background

Clinical practice universally assumes that appropriate empirical antibiotic therapy improves survival in patients with bloodstream infection. However, this is not generally supported by previous studies. We examined the association between appropriate therapy and 30-day mortality, while minimizing bias due to confounding by indication.

Methods

We conducted a retrospective cohort study between 2012 and 2017 at a tertiary university hospital in the Netherlands. Adult patients with bloodstream infection attending the emergency department were included. Based on in vitro susceptibility, antibiotic therapy was scored as appropriate or inappropriate. Primary outcome was 30-day mortality. To control for confounding, we performed conventional multivariable logistic regression and propensity score methods. Additionally, we performed an analysis in a more homogeneous subgroup (i.e. antibiotic monotherapy).

Results

We included 1,039 patients, 729 (70.2%) received appropriate therapy. Overall 30-day mortality was 10.4%. Appropriately treated patients had more unfavourable characteristics, indicating more severe illness. Despite adjustments, we found no association between appropriate therapy and mortality. For the antibiotic monotherapy subgroup (n = 449), patient characteristics were more homogeneous. Within this subgroup, appropriate therapy was associated with lower mortality (Odds Ratios [95% Confidence Intervals] ranging from: 0.31 [0.14 - 0.67] to 0.40 [0.19 - 0.85]).

Conclusions

Comparing heterogeneous treatment groups distorts associations despite use of common methods to prevent bias. Consequently, conclusions of such observational studies should be interpreted with care. If possible, future investigators should use our method of attempting to identify and analyse the most homogeneous treatment groups nested within their study objective, because this minimizes residual confounding.

INTRODUCTION

Bacterial infections can result in considerable mortality and have a profound global burden[203-205]. Patients with a severe infection (e.g. sepsis) often present in an acute care setting, such as the emergency department (ED). Initiation of targeted antibiotic therapy in the ED is important in patients with a suspected bacterial infection and is possible when the causative pathogen is proven by cultures with determination of the antibiogram[206]. However, this process usually takes over 24 hours and therefore empirical therapy is initiated in the ED. Appropriate empirical antibiotic therapy (i.e. appropriate therapy) is defined as applying the antibiotic agent which matches in vitro susceptibility of the isolated bacteria, but was initially provided without evidence on the causative pathogen or its antibiogram[59]. Clinical practice universally assumes that appropriate therapy improves survival in patients with bloodstream infection (BSI).

Although an overall beneficial outcome of appropriate antibiotic therapy in patients with BSI was demonstrated by meta-analyses[209, 210], studies that did not find lower mortality continued to be published[59, 243-249]. An explanation for these conflicting data is confounding by indication[250], yet this was not investigated in these studies[59, 243-249]. Confounding by indication arises because patients at risk of dying of BSI are more likely to receive broad spectrum antibiotic therapy - thus more often appropriate - as physicians want to ensure appropriateness most in severely ill patients[204]. This results in an imbalance in - measured and unmeasured - patient characteristics (i.e. underlying risk profile) between appropriately and inappropriately treated patients, thereby biasing the genuine relation between appropriate therapy and mortality[251].

The main objective of this study was to examine whether administration of appropriate empirical antibiotic therapy affects 30-day mortality in adult patients with BSI attending the ED, while minimizing bias due to confounding by indication. Subsequently, we focused on methodologically explaining why prior investigators suggested no impact of appropriate therapy on survival.

MATERIALS AND METHODS

Study design and setting

We conducted a retrospective cohort study at the Erasmus University Medical Center Rotterdam (Erasmus MC), which is a tertiary university hospital in the Netherlands. We used data from all patients attending the ED with BSI from July 2012 through December 2017. Blood cultures are taken in patients suspected for BSI, and subsequently empiric

antibiotic therapy is started. Antibiotic advice is protocolized in guidelines based on local and national prevalence and resistance data[222, 252]. These guidelines provide an advice depending on the suspected source of infection and clinical judgement of severity of disease, e.g. working diagnosis. The Medical Ethics Committee of the Erasmus MC reviewed our study and concluded that it did not fall under the scope of the Medical Research Involving Human Subjects Act and therefore no informed consent needed to be obtained. Our study is thus approved and registered under MEC-2018-1450.

Selection of participants

Patients were eligible for inclusion if they were at least 18 years of age and had a laboratory proven bacterial BSI at the ED. BSI was defined as presence of a known pathogen in one blood culture or a common commensal (e.g. Staphylococcus epidermidis)[214] in at least two blood cultures collected on separate occasions within two days from ED admission[214, 215]. Only the first episode of BSI was included to prevent domination of results by individuals that frequently visited the ED.

Data collection and processing

We combined electronic databases with data from the ED and the department of Medical Microbiology and Infectious Diseases. The ED database included empiric antibiotic therapy administered during the ED visit, potentially relevant and retrospectively available patient characteristics (serving as proxies for severity of disease), and mortality. Treatment strategy was either no antibiotic therapy, antibiotic monotherapy (if only one drug was administered), or antibiotic combination therapy (if more than one drug was administered). Also, patient charts were reviewed to assess dosage errors. General and demographic patients characteristics collected were: sex, age, arrival (by ambulance or other mode of transportation), triage category (according to the Manchester Triage System)[216], disposition (direct intensive care unit admittance or other), chills[217], vomiting[217], need for vasopressors, suspected site of infection (unknown, respiratory, abdominal, urogenital, skin or soft tissue, intravascular or thorax, central nervous system, other), and origin of infection (nosocomial or community-acquired)[218]. To account for severity of disease we used the first recorded vital signs (i.e. body temperature, heart rate, respiratory rate, systolic blood pressure, oxygen saturation, and consciousness), whether there was need for any supplemental oxygen, and calculated the National early warning score (NEWS)[219, 232]. Additionally, to account for comorbidity we collected all components of the age-adjusted Charlson comorbidity index (CCI)[221]. The primary outcome was 30-day mortality, because we expected 30 days to be a biologically plausible window to represent the effect of appropriate therapy on mortality[250]. For mortality data we used municipal death registration records.

The Medical Microbiology and Infectious Diseases database contained data about type of pathogen and their susceptibility (antibiogram) for all positive blood cultures collected at the ED. Blood cultures were performed using the BACTEC system (Becton Dickinson Diagnostic Instrument Systems, Sparks, Md) according to the manufacturers protocol. Type of pathogen was identified directly in one milliliter of blood by MALDI-TOF MS analysis (Microflex, Bruker Daltonics, Bremen, Germany). The in vitro susceptibility to antibiotic agents testing was performed with VITEK 2 (bioMérieux, Marcy l'Etoile, France). Based on earlier applied antibiotic therapy during ED visit and the established susceptibility of the isolated pathogen, we retrospectively determined the appropriateness of empirical therapy. In accordance with previous studies, no empiric antibiotic therapy, ineffective antibiotic therapy (based on antibiogram or if a dosage error was reported), or not intravenously administered antibiotic therapy (except for antibiotics with high bioavailability, i.e. metronidazole and ciprofloxacin) were all considered inappropriate. The interval of antibiotic administration was adjusted in patients with a glomerular filtration rate less than 30 mL per minute, however, this does not affect the initial dosage of antibiotic therapy administered in the ED[59, 209, 210, 244-250].

Data analysis and control for confounding bias

For descriptive statistics we examined all patient characteristics among appropriately versus (vs.) inappropriately treated patients. Based on distribution data were tested with an unpaired t-test, chi-squared test, or Fisher's exact test.

We considered patient characteristics as confounders during further analyses if, based on expert knowledge, controlling for the variable would reduce bias when studying the relation between appropriate therapy and 30-day mortality[251]. To improve our propensity score methods, we only included potential confounding variables in our models that were statistically related to outcome, as this decreases variance without increasing bias (*Appendix 10.A*)[253].

We conducted inferential statistics to investigate the association between appropriate therapy and 30-day mortality while attempting to control for confounding by indication. Results were presented as odds ratios (OR) with 95% confidence intervals (CI). We handled missing data using multiple imputations. For efficiency purposes we imputed 20 datasets using the chained equations method[254].

To limit confounding by indication, we controlled for measured proxies of disease severity (e.g. arrival mode, triage category, direct intensive care unit admittance, components of NEWS, components of CCI) with multiple statistical techniques. First, we performed a conventional multivariable logistic regression analysis. However, this method is

known to fall short in case of confounding by indication[255]. Therefore, secondly, we used propensity score methods. Propensity score methods directly focus on indication for treatment under study and potentially provide more precise estimates in studies in which confounding by indication may occur[255]. We applied three analytical procedures with the obtained propensity scores, namely 1) adjustment by logistic regression, 2) stratification, and 3) inverse probability of treatment weighting (*Appendix 10.A*)[256-258]. To assess the impact of potential contaminated BSI (i.e. those with a common commensal on multiple blood cultures), we subsequently performed a sensitivity analysis after exclusion of these patients.

Finally, we attempted to limit confounding bias by selecting patients treated with - appropriate or inappropriate - antibiotic monotherapy. When comparing the total appropriately to inappropriately treated group, we expected various degrees of confounding bias for different treatment strategies (i.e. no antibiotic therapy, antibiotic combination therapy, antibiotic monotherapy). We expected that patients with the lowest disease severity and the lowest risk of dying would more often receive no - thus inappropriate - antibiotic therapy. We also expected that severely ill patients with high chance of dying are more likely to receive antibiotic combination therapy to broaden the spectrum, resulting in more often appropriate therapy. Therefore, when studying the relation between appropriate therapy and mortality in the total population, including these treatment strategies potentially contributes to large heterogeneity between appropriately and inappropriately treated patients, which increases risk of confounding bias. We expected that the subset of patients who received antibiotic monotherapy was the least confounded group with more homogeneous measured and unmeasured confounders.

All hypothesis tests were 2-sided, with a significance level of $p < .05$. Statistical analyses were performed using R version 3.4.4.

RESULTS

Patient characteristics

We identified 1.286 adult patients with a positive laboratory proven blood culture taken at the ED. We excluded 247 patients with recurrent BSI, resulting in 1.039 unique patients of whom 729 (70.2%) received appropriate therapy. In 310 patients therapy was inappropriate: 184 patients received no empiric antibiotic therapy, 115 patients were treated with ineffective antibiotic therapy, and in 11 patients antibiotic therapy was not intravenously administered. Of the patients who were appropriately treated, cefuroxime and gentamicin combination therapy was most often administered. 30-day mortality was

10.4%. We found that 673 (64.8%) patients had a gram-negative BSI. The most frequently isolated pathogens were Escherichia coli (32.8%), Staphylococcus aureus (10.1%), and Streptococcus pneumoniae (8.2%). Patient characteristics are shown in *Table 10.1*.

Table 10.1: Patient characteristics in appropriately versus inappropriately treated patients (total population)

Characteristic	Appropriate n = 729 (70.2)	Inappropriate n = 310 (29.8)	p-value
Sex, male	425 (58.3)	201 (64.8)	.06
Age, mean (SD), years[A]	60.9 (15.5)	60.1 (15.9)	.44
Arrival by ambulance[A]	202 (27.7)	47 (15.2)	< .001
Triage category, acute/highly urgent[A,B]	205 (29.6)	33 (11.1)	< .001
Direct intensive care unit admittance[A]	66 (9.1)	8 (2.6)	< .001
Chills[A]	311 (42.7)	134 (43.2)	.92
Vomiting	178 (24.4)	68 (21.9)	.43
Need for vasopressors[A]	36 (4.9)	5 (1.6)	.02
Suspected site of infection, unknown	169 (23.2)	70 (22.6)	.90
Origin, nosocomial	384 (52.7)	175 (56.5)	.29
Antibiotic treatment strategy			
Combination therapy	382 (52.4)	22 (7.1)	< .001
Monotherapy	347 (47.6)	102 (32.9)	< .001
No antibiotic therapy	0 (0.0)	186 (60.0)	< .001
Vital signs / NEWS parameters			
Body temperature, mean (SD), °C[A,C]	38.4 (1.2)	38.0 (1.1)	< .001
Heart rate, mean (SD), /min[D]	108 (23.8)	100 (19.6)	< .001
Respiratory rate, mean (SD), /min[A,E]	24 (8.5)	21 (7.1)	< .001
Systolic blood pressure, mean (SD), mm Hg[A,F]	125 (28.5)	125 (24.5)	.77
Oxygen saturation, mean (SD), %[G]	95 (5.8)	96 (2.4)	< .001
Any supplemental oxygen[A]	339 (46.5)	62 (20.0)	< .001
Consciousness, not alert[A,H]	96 (15.5)	16 (6.5)	< .001
NEWS, mean (SD)	6.0 (3.8)	3.8 (3.1)	< .001
Comorbidities of Charlson comorbidity index[I]			
Diabetes mellitus, uncomplicated	147 (20.2)	53 (17.1)	.29
Diabetes mellitus, end-organ damage[A]	10 (1.4)	3 (1.0)	.77
Liver disease, mild[A]	93 (12.8)	47 (15.2)	. 35
Malignancy, leukemia, lymphoma, localized solid tumor[A]	120 (16.5)	61 (19.7)	.25
Malignancy, metastatic solid tumor[A]	93 (12.8)	40 (12.9)	> .99
Chronic kidney disease[A]	124 (17.0)	45 (14.5)	.37
Congestive heart failure	96 (13.2)	37 (11.9)	.66
Myocardial infarction	103 (14.1)	36 (11.6)	.32
Chronic obstructive pulmonary disease[A]	95 (13.0)	39 (12.6)	.92

Table 10.1: Patient characteristics in appropriately versus inappropriately treated patients (total population) (*continued*)

Characteristic	Appropriate n = 729 (70.2)	Inappropriate n = 310 (29.8)	p-value
Perivascular disease	77 (10.6)	44 (14.2)	.12
Cerebrovascular accident or transient ischemic attack[A]	115 (15.8)	26 (8.4)	.002
Dementia[A]	30 (4.1)	6 (1.9)	.12
Connective tissue disease	57 (7.8)	20 (6.5)	.52
Peptic ulcer disease	17 (2.3)	8 (2.6)	> .99
Type of isolated pathogen			
Gram-negative BSI	457 (62.7)	216 (69.7)	.03

Data are presented as N (%) unless otherwise indicated. Data in this table is not imputed yet. Abbreviations: NEWS, national early warning score; BSI, bloodstream infection.
[A]Confounding variables.
[B]Data on triage category were missing for 50 (4.6%) patients.
[C]Data on body temperature were missing for 9 (0.9%) patients.
[D]Data on heart rate were missing for 24 (2.3%) patients.
[E]Data on respiratory rate were missing for 370 (35.5%) patients.
[F]Data on systolic blood pressure were missing for 20 (1.9%) patients.
[G]Data on oxygen saturation were missing for 43 (4.3%) patients.
[H]Data on consciousness were missing for 175 (16.8%) patients.
[I]Comorbidities with a prevalence below 1% are not presented (i.e. moderate to severe liver disease, acquired immunodeficiency syndrome, and hemiplegia).

Patients receiving appropriate therapy had less favourable measured characteristics than patients receiving inappropriate antibiotic therapy: they more frequently arrived by ambulance (27.7% vs. 15.2%), had higher triage categories (29.6 % vs. 11.1%), were more often admitted directly to the intensive care unit (9.1% vs. 2.6%), needed vasopressors more frequently (4.9% vs. 1.6%), and received more antibiotic combination therapy (52.4% vs. 7.1%). In addition, appropriately treated patients had more abnormal vital signs and on average a higher NEWS of 6.0 (± 3.8) vs. 3.8 (± 3.1).

Appropriate empirical antibiotic therapy and 30-day mortality

Crude 30-day mortality for appropriately treated patients was 11.1% (81 patients) vs. 8.7% (27 patients) for inappropriately treated patients (OR [95%CI]: 1.31 [0.84 - 2.10]). There was no association between appropriate therapy and 30-day mortality after conventional adjustment for confounders, adjustment for propensity score, propensity score stratification and inverse probability of treatment weighting (OR [95%CI] ranging from: 0.71 [0.43 - 1.19] to 1.03 [0.76 - 1.40], *Figure 10.1*).

For sensitivity analysis, we examined the impact of excluding patients with common commensal bacteria on multiple blood cultures collected on separate occasions within two days from ED admission. In our study, 24 patients had at least two subsequent blood cultures with a common commensal (17 Staphylococcus epidermidis, 3 Staphylococcus

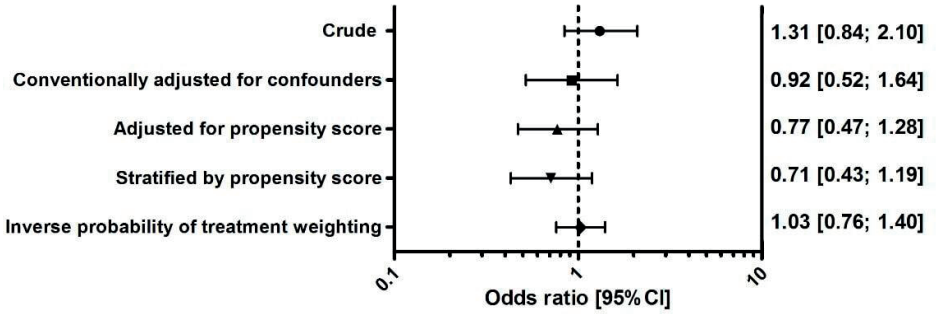


Figure 10.1: Estimates of appropriate empirical antibiotic therapy on 30-day mortality (total population).
Abbreviations: CI, confidence interval.
Confounding variables: age, arrival, triage category, direct intensive care unit admittance, chills, need for vasopressors, body temperature, respiratory rate, systolic blood pressure, supplemental oxygen, consciousness, diabetes mellitus with end-organ damage, mild liver disease, malignancy, chronic kidney disease, chronic obstructive pulmonary disease, cerebrovascular accident or transient ischemic attack, and dementia.
For a detailed description of statistical adjustment techniques, see Appendix 10.A.
This figure shows attenuation of estimates after adjustment for confounders.

hominis, 1 Bacillus licheniformis, 1 Rhodococcus equi, 1 Staphylococcus capitis, and 1 Staphylococcus lugdunensis). Appropriate therapy was administered in 9 (37.5%) patients. Excluding these patients did not affect our results.

Subgroup analysis antibiotic monotherapy

There were 449 patients treated with antibiotic monotherapy of whom 347 (77.3%) received appropriate therapy. In 102 patients therapy was inappropriate: 92 patients were treated with ineffective antibiotic therapy and in 10 patients antibiotic therapy was not intravenously administered. Of the patients who were appropriately treated, cefuroxime was most often administered. 30-day mortality was 7.1%. We found that 299 (66.6%) patients had a gram-negative BSI, which is comparable to the rate of gram-negative BSI in the total population (64.8%). The most frequently isolated pathogens were Escherichia coli (35.4%), Staphylococcus aureus (11.1%), and Klebsiella pneumoniae (7.8%). Patient characteristics were comparable for appropriately and inappropriately treated patients, indicating more homogeneity in the monotherapy subgroup compared to the total population (*Table 10.2*).

In the monotherapy subgroup, crude 30-day mortality for appropriately treated patients was 5.5% (19 patients) vs. 12.7% (13 patients) for inappropriately treated patients. Appropriate therapy was associated with lower 30-day mortality after crude estimation, adjustment for propensity score, propensity score stratification, and inverse probability of treatment weighting (OR [95%CI] ranging from: 0.31 [0.14 - 0.67] to 0.40 [0.19 - 0.85], *Figure 10.2*). Conventional adjustment for confounders had an OR with 95%CI of 0.41 [0.14 - 1.18].

Table 10.2: Patient characteristics in appropriately versus inappropriately treated patients (antibiotic monotherapy)

Characteristic	Appropriate n = 347 (77.3)	Inappropriate n = 102 (22.7)	p-value
Sex, male	200 (57.6)	67 (65.7)	.18
Age, mean (SD), years[A]	60.1 (15.4)	63.0 (15.1)	.09
Arrival by ambulance[A]	55 (15.9)	14 (13.7)	.71
Triage category, acute/highly urgent[A]	52 (15.7)	12 (12.2)	.49
Direct intensive care unit admittance[A]	10 (2.9)	2 (1.9)	> .99
Chills[A]	164 (47.3)	47 (46.1)	.92
Vomiting	86 (24.8)	21 (20.6)	.46
Need for vasopressors[A]	3 (0.9)	2 (2.0)	.70
Suspected site of infection, unknown	86 (24.8)	20 (19.6)	.34
Origin, nosocomial	207 (59.7)	63 (61.8)	.79
Vital signs / NEWS parameters			
Body temperature, mean (SD), °C[A]	38.3 (1.1)	38.1 (1.2)	.05
Heart rate, mean (SD), beats/min	103 (20.6)	100 (21.6)	.21
Respiratory rate, mean (SD), breaths/min[A]	21 (7.0)	20 (6.4)	.21
Systolic blood pressure, mean (SD), mm Hg[A]	128 (25.7)	123 (21.1)	.05
Oxygen saturation, mean (SD), %	96 (5.5)	96 (2.3)	.67
Any supplemental oxygen[A]	106 (30.5)	33 (32.4)	.82
Consciousness, not alert[A]	18 (6.3)	7 (8.5)	.65
NEWS, mean (SD)	4.5 (3.0)	4.3 (3.4)	.48
Comorbidities of Charlson comorbidity index[B]			
Diabetes mellitus, uncomplicated	64 (18.4)	17 (16.7)	.79
Diabetes mellitus, end-organ damage[A]	5 (1.4)	0 (0.0)	.59
Liver disease, mild[A]	53 (15.3)	15 (14.7)	> .99
Malignancy, leukemia, lymphoma, localized solid tumor[A]	64 (18.4)	20 (19.6)	.90
Malignancy, metastatic solid tumor[A]	45 (13.0)	19 (18.6)	.20
Chronic kidney disease[A]	85 (24.5)	21 (20.6)	.49
Congestive heart failure	52 (15.0)	11 (10.8)	.36
Myocardial infarction	48 (13.8)	16 (15.7)	.76
Chronic obstructive pulmonary disease[A]	39 (11.2)	19 (18.6)	.07
Perivascular disease	31 (8.9)	13 (12.7)	.34
Cerebrovascular accident or transient ischemic attack[A]	57 (16.4)	11 (10.8)	.21
Dementia[A]	11 (3.2)	1 (1.0)	.39
Connective tissue disease	27 (7.8)	6 (5.9)	.67
Peptic ulcer disease	9 (2.6)	3 (2.9)	.88
Type of isolated pathogen			
Gram-negative BSI	214 (61.7)	85 (83.3)	< .001

Data are presented as N (%) unless otherwise indicated. Data in this table is not imputed yet.
Abbreviations: NEWS, national early warning score; BSI, bloodstream infection.
[A]Confounding variables.
[B] Comorbidities with a prevalence below 1% are not presented (i.e. moderate to severe liver disease, acquired immunodeficiency syndrome, and hemiplegia).

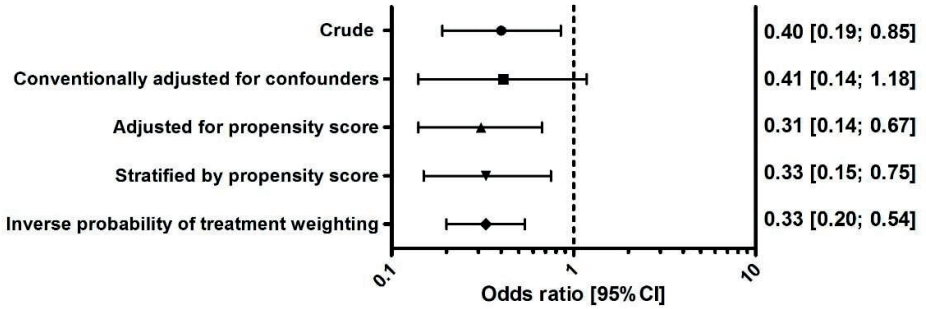


Figure 10.2: Estimates of appropriate empirical antibiotic therapy on 30-day mortality (antibiotic monotherapy).
Abbreviations: CI, confidence interval.
Confounding variables: age, arrival, triage category, direct intensive care unit admittance, chills, need for vasopressors, body temperature, respiratory rate, systolic blood pressure, supplemental oxygen, consciousness, diabetes mellitus with end-organ damage, mild liver disease, malignancy, chronic kidney disease, chronic obstructive pulmonary disease, cerebrovascular accident or transient ischemic attack, and dementia.
For a detailed description of statistical adjustment techniques, see Appendix 10.A.
This figure shows attenuation of estimates after adjustment for confounders.

DISCUSSION

This study aimed to address the confounding that exists in establishing the effects of antibiotic appropriateness in patients with BSI. Despite extensive adjustment for confounding, we found no association between appropriate empirical antibiotic therapy and mortality when assessing all patients. This finding - in line with previous studies[59, 244-249] - remains counterintuitive and is in contrast to fundamentals of current clinical practice[204].

We hypothesized that confounding by indication was the explanation for finding no association between appropriate therapy and mortality in previous studies. Patients at risk of dying of BSI are more likely to receive broad spectrum antibiotic - thus more often appropriate - therapy as physicians want to ensure appropriateness most in severely ill patients. As a result, the association between appropriate therapy and mortality is biased. In our study, the first clue for confounding by indication was more unfavourable patient characteristics in the appropriately treated group. We noticed this heterogeneity as well in the study of *Anderson et al.*, which also found no association between appropriate therapy and mortality[244]. However, the authors did not consider confounding by indication as a potential explanation for their findings[244]. A second clue for confounding was attenuation of estimates when controlling for bias - with both conventional multivariable logistic regression and propensity score methods. We noticed that in prior studies, that also found no association, there was attenuation of estimates after adjustment

for confounders[248, 249]. Since we only adjusted for observed confounders, unmeasured - residual - confounders could still be of potential bias.

Chance of residual confounding is absent in totally homogenous groups (e.g. as in an ideal randomized controlled trial)[251]. Our total population was heterogeneous in measured patient characteristics and we expected various degrees of confounding bias for different treatment strategies. We expected that patients receiving antibiotic combination - thus more often appropriate - therapy were the most ill and patients receiving no antibiotic therapy - thus inappropriate therapy - were the least ill patients. We expected the remainder of patients that received antibiotic monotherapy to be more comparable, as physicians chose to treat these patients presumably based on a more comparable judgment of illness. In addition, the severely confounded treatment strategies - i.e. antibiotic combination therapy and no antibiotic therapy - are per definition excluded during this subgroup analysis. We therefore decided to subsequently analyse the antibiotic monotherapy subgroup. We found that for antibiotic monotherapy measured patient characteristics of appropriately and inappropriately treated patients were more balanced (i.e. homogeneous), lowering the chance of residual confounding. In this subgroup appropriate therapy was associated with lower 30-day mortality. This finding is in line with our expectations and current practice, and supports our hypothesis that residual confounding distorts associations when comparing heterogeneous treatment groups.

Reducing confounding by indication through analysing homogeneous subgroups - in our study antibiotic monotherapy - is not often done. Previous studies on appropriate therapy and mortality disregarded severely confounded treatment strategies (i.e. antibiotic combination therapy, no antibiotic therapy), which resulted in comparison of heterogeneous groups[59, 244-249]. Therefore, the conclusions of these studies are potentially not trustworthy.

To prevent confounding, we adjusted for validated risk scores (e.g. NEWS, CCI) and applied several adjustment techniques (i.e. conventional multivariable logistic regression and propensity score methods). However, for the total population, these techniques fell short and we were unable to prevent bias. Apparently, a physicians' decision to initiate a certain therapy is not only based on findings that are represented by such risk score systems, hence statistical adjustment techniques fall short. Thus, conclusions of observational studies comparing heterogeneous groups should be interpreted with care. If possible, future investigators should use our method of attempting to identify and analyse the most homogeneous treatment groups nested within their study objective, as we demonstrated that this minimizes residual confounding.

Limitations

Our study has limitations. First, we used retrospectively collected data making our study prone to bias. However, the quality of available data was assumed to be high as all data used was essential for daily clinical practice. For only 13 patients (1.3%) documentation was unclear on whether antibiotic therapy was administered at the ED or after discharge, therefore we scored them as inappropriate therapy. We had no data on the exact time to the first antibiotic dose, but only on whether administration was during the ED visit or not. However, timing of antibiotic administration would have had no impact on the outcome of the inappropriately treated group. Delayed treatment might have had an impact on mortality in the appropriately treated group, which could have led to a more extreme estimate than we found already.

Furthermore, we want to emphasize that we considered the association between empiric antibiotic treatment at the ED and 30-day mortality, as this was our main study objective. Depending on disease course and culture results, antibiotic treatment could have been modified later on resulting in a different definitive antibiotic treatment. Also, we had no data on whether any source control such as abscess drainage was performed after ED discharge. Aside from empiric antibiotic treatment at the ED, this may have altered survival as well.

CONCLUSIONS

We initially found that appropriate empirical antibiotic therapy was not beneficial in patients with BSI. We showed that this counterintuitive finding was presumably the result of residual confounding. Patients that present with high disease severity are more likely to receive appropriate therapy than less ill patients. Therefore, the appropriately treated are initially at higher risk of dying than the inappropriately treated. Analysing these heterogeneous treatment groups results in distorted associations and subsequent conclusions despite the use of common methods to prevent bias. With a subgroup analysis in a more homogeneous population (i.e. antibiotic monotherapy), we found the expected benefit of appropriate therapy. Our study underlines the complexities of performing clinical observational research. In case of heterogeneous groups results should always be interpreted with care. If possible, future investigators should attempt to identify and analyse the most homogeneous treatment groups nested within their study objective, because this minimizes residual confounding.

Appendix 10.A: Propensity score methods

We obtained the propensity score by multivariable logistic regression, with appropriate empirical antibiotic therapy (AEAT) as dependent variable and all confounders as independent variables. The propensity score is a balancing score, ranging from 0 to 1, representing probability of AEAT assignment conditional on observed confounders[256].

We applied three analytical procedures with the obtained propensity score. First, we used the propensity score as single independent covariate representing all confounders during logistic regression. Then we stratified on propensity score by bins of 0.1. For patients within the same bin, distribution of observed confounders is conditionally similar for appropriately and inappropriately treated patients if there is overlap in propensity score. This concept mimics process of randomization. After trimming all patients with non-overlapping propensity scores we obtained odds ratios with standard comparison and performed Mantel-Haenszel pooling[256]. Finally, we used inverse probability of treatment weighting as adjustment technique, which uses the propensity score as a weight during subsequent standard comparison[253].

Based on previous simulation studies, we only included potential confounding variables in our statistical models that were statistically related to outcome (relative risk > 1.3) as this decreases variance without increasing bias[253]. This is mainly important for our propensity score model. Including variables not associated with outcome (30-day mortality), but with exposure (AEAT) can lead to overseparation.

Propensity scores were estimated in our total population and subsequently used in subgroup analyses (i.e. antibiotic monotherapy). Recent simulation studies showed this is a feasible approach[258].

chapter 11

Drug Non-adherence is a Common But Often Overlooked Cause of Hypertensive Urgency and Emergency at the Emergency Department

Naomi Overgaauw
Jelmer Alsma
Anniek Brink
Edon Hameli
Soma Bahmany
Laura E.J. Peeters
Anton H. van den Meiracker
Stephanie C.E. Schuit
Birgit C.P. Koch
Jorie Versmissen

J Hypertens. 2019 May;37(5):1048-1057.

ABSTRACT

Background

Over 70% of patients who visit the emergency department (ED) with a hypertensive emergency (HE) or a hypertensive urgency (HU) have previously been diagnosed with hypertension. Drug non-adherence is assumed to play an important role in development of HU and HE, but exact numbers are lacking. We aimed to [1] retrospectively compare characteristics of patients with HU and HE and [2] to prospectively quantify the attribution of drug non-adherence.

Methods

[1] We retrospectively analysed clinical data including information on non-adherence obtained by treating physicians of patients with systolic blood pressure (SBP) ≥ 180 mmHg and diastolic blood pressure (DBP) ≥ 110 mmHg visiting the ED between 2012-2015. [2] We prospectively studied drug adherence among patients admitted to the ED with severely elevated BP by measuring plasma drug levels using liquid chromatography tandem mass spectrometry from September 2016 to March 2017.

Results

[1] Of the 1163 patients retrospectively analysed, 257 (22.0%) met the criteria for HU and 356 (30.6%) for HE. Mean SBP (SD) was 203 (19) mmHg and mean DBP 121 (12) mmHg. Mean age was 60.1 (14.6) years; 55.1% were male. In 6.3% of patients with HU or HE non-adherence was recorded as an attributing factor. [2] Of the 59 patients prospectively analysed, 18 (30.5%) were non-adherent for at least one of the prescribed antihypertensive drugs.

Conclusions

HU and HE are common health problems resulting in frequent ED admissions. Workup of patients with a HU or HE should include an assessment of drug adherence to optimize treatment strategy.

Keywords

Antihypertensive agents, Medication adherence, Emergency care

INTRODUCTION

A markedly elevated blood pressure (BP) is a common finding at the emergency department (ED): at least 5 percent of patients in the ED have one or more severely elevated measurements, usually defined as systolic BP ≥ 180 mmHg or diastolic BP ≥ 120 or 110 mmHg, although terminology and cut-offs differ between studies[259, 260]. In most of these patients the high BP is transient and is a reaction to pain, anxiety or stress. This is sometimes referred to as a hypertensive pseudo-crisis, and warrants no further action[261]. Around 0.5 percent of ED visits are primarily for severe hypertension. In such cases, the most important aim is to differentiate between a "hypertensive emergency" (HE), when acute target organ damage is present or impending, and a "hypertensive urgency" (HU) when this is not the case[60, 65, 262]. HU and HE were previously summarized as "hypertensive crisis" but since this terminology seems outdated we will use the terms HU and HE[263, 264]. HE requires immediate action to lower the BP using intravenous antihypertensive drugs in an intensive or high care unit, while HU allows BP regulation using oral therapy in an outpatient setting[60, 260]. When patients visit the ED primarily for severe hypertension, depending on complaints and findings of a physical examination, extensive tests (e.g. laboratory testing, ophthalmoscopy) may be needed to distinguish between HU and HE and to determine whether hospital admission is necessary[65, 265].

HU and HE can occur in patients with previously unidentified hypertension as a first presentation of their hypertensive condition. However, over 70% of patients presenting at the ED have been previously diagnosed with hypertension and have been prescribed antihypertensive drugs[260, 265-267]. Drug non-adherence, defined as not taking drugs as previously agreed on with the treating physician, is assumed to play an important role in the development of a HE and HU, but exact numbers are lacking. Poor drug adherence of antihypertensive and other cardiovascular drugs is associated with a higher risk of developing cardiovascular disease[268, 269].

When a patient presents at the ED with severe hypertension, it is crucial to distinguish non-adherence to therapy from treatment failure. In non-adherent patients, physicians should discuss reasons for non-adherence and methods to improve adherence, whereas in adherent patients drug therapy should be optimized.

In this study, we combined a retrospective and a prospective study to answer two related and important research questions considering severely elevated BP at the ED. The first objective was to compare characteristics of patients with HU and those with HE, including assessment of drug adherence by the treating physician. The second objective was to prospectively determine the incidence of non-adherence to prescribed

antihypertensive drugs in patients with severely elevated BP at the ED by measuring plasma drug levels.

METHODS

Study design

In this manuscript we describe two studies. [1] We performed a retrospective cross-sectional study among patients who visited the ED from January 1st 2012 to December 31st 2015 with at least one BP measurement. Due to the large number of patients with elevated BP caused by stress, anxiety or pain, we restricted the number of cases for analysis by choosing to only include patients who met both the SBP and DBP cut-off values. [2] We performed a prospective study in which we analysed plasma drug levels of prescribed antihypertensive drugs in patients who visited the ED from September 1st 2016 with severely elevated BP suspected of HU or HE. Here we used the formal cut-offs for HE described in the current American and European guidelines (SBP ≥ 180 mmHg or DBP ≥ 120 mmHg)[263].

Study population

The studies were performed at Erasmus University Medical Center in Rotterdam, The Netherlands (Erasmus MC), which is a large urban tertiary care hospital. The ED is an open access department located in the city centre, and has visits from approximately 30,000 patients annually.

Retrospective study

We used a database containing all patient records from ED visits in the period from January 1st 2012 to December 31st 2015 to select patients who had a SBP ≥ 180 mmHg and a DBP ≥ 110 mmHg at triage. Patients 18 years of age and older were included. For patients with multiple visits to the ED during the inclusion period, only the first visit was included.

Prospective study

Inclusion commenced from September 1st 2016 until the number of patients required was reached as determined in sample size calculations. We included all patients aged 18 years or older presenting to the ED or the fast-track program with a SBP ≥ 180 mmHg or a DBP ≥ 120 mmHg at triage, who were prescribed one or more antihypertensive drugs that we were able to measure in plasma at least 24 hours after intake using a validated liquid chromatography-tandem mass spectrometry (LC-MS/MS), and from whom routine blood samples were obtained[270]. We excluded patients who were unable to give informed consent or when severe hypertension was likely to have been caused by severe pain or stress.

Variables and measurement

We defined HE as a severely elevated BP according to the inclusion criteria with the presence of acute end-organ damage (i.e. ischemic stroke, haemorrhagic stroke, myocardial infarction, unstable angina, acute aortic dissection, acute pulmonary oedema, hypertensive encephalopathy and bilateral hypertensive retinopathy grade 3 or 4[261, 264]. HU was defined as severely elevated BP without acute or impending end-organ damage[260, 261, 264]. Patients were labelled as "non-HE and non-HU severe hypertension" when the BP was a result of extreme pain, anxiety or stress. This was based upon reasons for referral or presentation (other than hypertension) to the ED, on physicians' remarks in patient files and spontaneous recovery of BP after pain or stress relief.

We manually extracted data from electronic patient records including demographic data (i.e. age, gender), complaints (specifically headache, distorted vision, chest pain, palpitations, paraesthesia, paresis, gastrointestinal complaints, pain at any location), medical history, information on use of drugs and on drugs of abuse. When available, we collected test results including laboratory measurements, electrocardiograms (ECGs) focusing on left ventricular hypertrophy (LVH) using Sokolow-Lyons criteria[271] and radiological examinations (i.e. chest radiography for cardio-thoracic ratio (CTR) assessment: > 0.5 was considered enlarged). The working diagnosis and patient disposition after discharge from the ED were recorded.

Measuring drug levels and definition of non-adherence

All patients received standard care in the ED. In this workup, routine blood samples were taken to diagnose or exclude end-organ damage (e.g. measurement of serum creatinin level, presence of fragmentocytes)[264]. For the prospective study, we used the remainder of these blood samples to measure levels of prescribed antihypertensive drugs in plasma using a validated LC-MS/MS multimethod[270]. Using this method we were able to detect losartan, valsartan, enalapril, perindopril, spironolactone, amlodipine and nifedipine and four active metabolites perindoprilate, enalaprilate, losartan carboxylic acid and canrenone. With this method drug levels are detectable for 24 hours or more after intake, allowing an objective assessment on adherence without knowledge of last moment of drug intake. Partial non-adherence was defined as self-reported non-adherence or non-detectable (concentration < lower limit of detection) drug levels of one of the prescribed drugs or its active metabolite, and complete non-adherence as self-reported non-adherence to all drugs or non-detectable drug levels of all tested drugs. Drug levels exactly at the lower level of detection, in other words extremely low drug levels, were scored based on time of last ingestion, drug levels of other antihypertensive drugs taken at the same time and discussion of (non-)adherence during stay in the hospital.

Statistical methods

Patient characteristics were presented as mean and standard deviation (SD), or as an absolute number (proportion). For the retrospective study, continuous variables were compared with one-way analysis of variance (ANOVA). Categorical variables were compared using the Pearson chi-squared test. For all variables for which > 10% was missing, "missingness" is shown separately. A p-value < 0.05 was considered significant. All analyses were conducted with IBM SPPS Statistics for Windows version 21 (IBM Corp., Armonk, N.Y., USA).

For the prospective study, we performed a pre-emptive power calculation based on assumptions, since this was the first study to assess non-adherence at the ED in patients with a suspicion of a HE. Assuming 50% non-adherence, 50 patients needed to be included to have 80% power (1-sample t-test).

Ethical approval

For the retrospective study, the Medical Ethics Committee of Erasmus MC concluded that the study did not fall under the scope of the Medical Research Involving Human Subjects Act (WMO), due to its retrospective nature and the anonymization of patient details, therefore no informed consent needed to be obtained. For the prospective study, the Medical Ethics Committee concluded that the study did not fall under the scope of the Medical Research Involving Human Subjects Act, since previously obtained blood samples were used. Informed consent was only deemed necessary for the collection and analyses of clinical data. Patients were informed about the study and informed consent was obtained from all eligible patients, and a withdrawal of consent form was given due to the short consideration time. To prevent a potential bias assuming non-adherent patients are less likely to give consent, we anonymously analysed samples of patients who did not give or withdrew consent after registering the expected drugs. For patients who gave consent data were collected from the electronic patient records.

RESULTS

Retrospective study

A total of 123,552 patients visited the ED in the 4-year inclusion period, of whom 64,979 had a recorded BP measurement. Of these ED visits, 1,237 (1.9%) had SBP ≥ 180 mmHg and DBP ≥ 110 mmHg. Since 75 patients visited the ED at least two times, we analysed a total of 1163 patients (*Figure 11.1*).

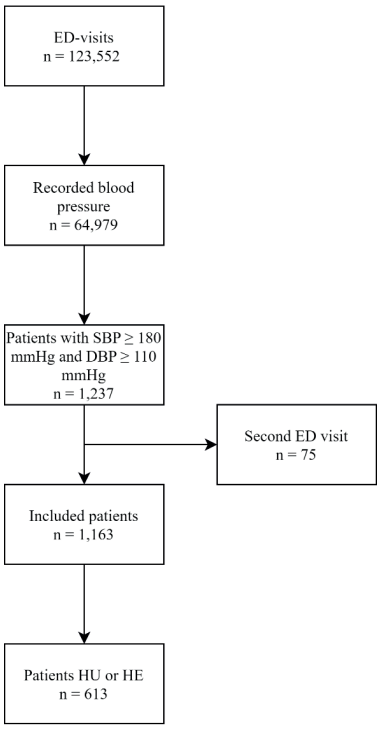


Figure 11.1: Flow chart retrospective study. Abbreviations: ED, emergency department; HU, hypertensive urgency; HE, hypertensive emergency

Incidence of patients visiting the ED with BP ≥ 180/110 mmHg, increased from 136 patients (0.6% of all ED visits) in 2012 to 414 patients (1.9%) in 2015. Patients were predominantly male with a mean age of 60 years (SD 15 years). Mean SBP was 203 mmHg (SD 19 mmHg) and mean DBP was 121 mmHg (SD 12 mmHg) (*Table 11.1*).

Of all patients presenting with severely elevated BP, the combined incidence of HU and HE was 52.7%, of which HE was diagnosed more frequently than HU (30.6% vs. 22.1 %; $p < 0.001$). Patients with HE were older than patients with HU (64 years vs. 58 years; $p < 0.001$) and had a higher BPs (SBP: 209 mmHg vs. 203 mmHg; $p < 0.001$ and DBP: 124 vs. 121 mmHg; $p < 0.001$). The most frequent diagnoses in patients with HE were stroke (10.7% ischemic, 8.5% haemorrhagic), acute pulmonary oedema (4.1%) and myocardial infarction (4.0%) (*Table 11.1*). Since women suspected of (pre)eclampsia were according to local guidelines referred to the obstetric clinic, no preeclampsia was recorded.

Seventy-nine patients (6.8%) were referred primarily for hypertension to exclude HE. Of these, 57 patients (72.2%) met the criteria of HU and seven (8.9%) of HE. The most

Table 11.1: Patient characteristics retrospective study

	Total (n = 1,163)	No HE / HU(e.g. severe pain; n = 550)	Hypertensive urgency (n = 257)	Hypertensive emergency (n = 356)
Mean (SD) age (years)	60.1 (14.6)	58.8 (14.9)	58.2 (13.4)	63.5 (14.7)
Male	641 (55.1)	301 (54.7)	125 (48.6)	215 (60.4)
Mean (SD) SBP (mmHg)	203 (19)	199 (16)	203 (17)	209 (20)
Mean (SD) DBP (mmHg)	121 (12)	119 (10)	121 (12)	124 (14)
Mean (SD) BMI (kg/m^2)	27 (5.6)	26.7(5.4)	28.0 (5.8)	27.3 (5.7)
History Hypertension	695 (59.8)	292 (53.0)	173 (67.3)	230 (64.4)
History Hypertensive crisis				
Yes	57 (4.9)	14 (2.5)	21 (8.2)	22 (6.2)
No	411 (35.3)	212 (38.5)	89 (34.6)	110 (30.9)
Missing	695 (59.8)	324 (58.9)	147 (57.2)	224 (62.9)
Diabetes	208 (17.9)	103 (18.7)	31 (12.1)	74 (21.2)
Alcohol use				
Yes	284 (24.4)	125 (23.3)	73 (28.4)	86 (24.3)
No	317 (27.3)	121 (22.0)	88 (34.2)	108 (30.3)
Missing	561 (58.3)	303 (55.3)	96 (37.4)	162 (45.5)
Smokers				
Yes	270 (23.2)	97 (17.7)	70 (27.2)	103 (28.9
No	454 (39.1)	186 (33.9)	124 (48.2)	144 (40.4)
Missing	438 (37.7)	266 (48.5)	63 (24.5)	109 (30.6)
Previously reported non-adherence				
Yes	111 (9.5)	35 (6.4)	38 (14.8)	38 (10.7)
No	355 (30.5)	167 (30.4)	92 (35.8)	96 (27.0)
Missing	697 (59.9)	348 (63.3)	127 (49.4)	222 (62.4)
Chronic kidney disease	92 (7.9)	41 (7.4)	18 (7.0)	33 (9.3)
Previous stroke or TIA	136 (11.7)	41 (7.4)	26 (10.2)	69 (19.5)
Previous coronary artery disease	110 (9.5)	39 (7.1)	29 (11.2)	42 (11.8)
Antihypertensive drugs prescribed	591 (50.8)	256 (50.6)	147 (57.2)	191 (53.6)
Number antihypertensive drugs				
Mean (SD) overall	1 (1.1)	1 (1.4)	1.1 (1.2)	1.2 (1.6)
Mean (SD) users only	2 (1.0)	1.7 (0.95)	2 (1.0)	2 (1)
Taking ≥ 2 antihypertensive drugs	319 (27.4)	126 (22.9)	87 (33.9)	106 (29.7)
Taking ≥ 3 antihypertensive drugs	150 (12.8)	59 (10.7)	40 (15.6)	51 (14.3)
Angiotensin II receptor blocker	129 (11.1)	55 (10.0)	29 (11.3)	45 (12.6)
ACE-inhibitor	217 (18.4)	91 (16.5)	51 (19.9)	75 (21.2)
Calcium channel blocker	169 (14.5)	79 (14.3)	44 (17.2)	46 (13.0)
Diuretic	186 (16.0)	76 (13.8)	46 (17.9)	64 (17.9)
Betablocker	304 (26.1)	117 (21.2)	80 (31.1)	107 (30.0)
Fixed-dose combination	57 (4.9)	26 (4.7)	15 (5.9)	16 (4.5)
Target organ damage				
Haemorrhagic Stroke	99 (8.5)		-	99 (8.5)
Ischaemic Stroke	124 (10.7)		-	124 (10.7)
Pulmonary oedema	48 (4.1)		-	48 (4.1)
Myocardial infarction	46 (4.0)		-	50 (4.0)
Aortic dissection	9 (0.8)		-	9 (0.8)
Acute kidney failure	4		-	4
Retinopathy grade III/IV	10		-	10
Thrombotic microangiopathy	1		-	1

Table 11.1: Patient characteristics retrospective study (*continued*)

	Total (n = 1,163)	No HE / HU(e.g. severe pain; n = 550)	Hypertensive urgency (n = 257)	Hypertensive emergency (n = 356)
Suspicion non-adherence				
Yes	54 (4.6)	15 (2.7)	24 (9.3)	15 (4.2)
Missing	1,108 (95.3)	535 (97.3)	233 (90.7)	341 (95.8)
Hospital admission	764 (65.6)	166 (51.7)	92 (35.6)	335 (93.8)
Mean (SD) days	6 (12)	4.9 (7.4)	2.1 (4)	8.3 (14)
ICU admission	103 (8.9)	23 (7.7)	0	61 (17.2)
Change in drug regime after discharge (ED/hospital)	388 (28.7)	80 (14.5)	115 (60.5)	191 (54.2)
Drugs first 24 hours				
Labetalol intravenously	145 (12.5)	18 (3.3)	8 (3.1)	119 (33.7)
Nitroglycerin intravenously	89 (7.7)	12 (2.2)	16 (6.3)	61 (17.1)
Nifedipine retard	121 (10.4)	27 (4.9)	69 (27.0)	25 (7.1)
Captopril	32 (2.8)	0	25 (9.8)	7 (2.0)
Died after admission	92 (8)			
Inhospital mortality	108 (9.2)	47 (5.3)	0	57 (16.2)

Values are numbers (percentages) unless stated otherwise. Abbreviations: HE, hypertensive emergency; HU, hypertensive urgency; SD, standard deviation; SBP, systolic blood pressure; DBP, diastolic blood pressure, BMI, body mass index, TIA, transient ischemic attack; ACE, angiotensin converting enzyme; ICU, intensive care unit.

frequently reported symptoms were headache, symptoms of the gastro-intestinal tract, chest pain and dyspnoea (*Table 11.2*).

Sixty-four patients with either HE or HU were asymptomatic. Of this group, 55 patients had HE (85.9%) and nine (14.1%) HU. In most cases (90.3%) laboratory tests were performed. An ECG was made for 65.3% of patients, which showed left ventricular hypertrophy in 19.1% of patients (*Table 11.3*). Of all patients, 572 (49.2%) patients were not prescribed any antihypertensive drugs, 272 (23.3%) patients were prescribed one antihypertensive drug and 319 (27.4%) patients were prescribed two or more. Of the patients using antihypertensive drugs 26.1% used beta blockers, 18.4% angiotensin converting enzyme (ACE)-inhibitors, 16% diuretics, 14.5% calcium channel blockers and 11.1% angiotensin-II receptor blockers. Only 4.9% of prescribed antihypertensive drugs were a fixed-dose combination.

Labetalol was used most often when treatment with an intravenous antihypertensive drug was needed and was given to 23.1% of all patients, and to 59.2% of patients with HE. The most commonly administered oral antihypertensive drug was nifedipine retard (10.4%). Oral therapy was mostly given in case of a HU, although a limited number of patients with a HE (9%) also received oral therapy. Suspicion of non-adherence was documented more often in patients with HU than those with HE (9.4% vs. 4.2%; $p < 0.001$).

Table 11.2: Symptoms retrospective study

Symptoms	Total (n = 1,163)	Not HE/HU (n = 550)	Hypertensive urgency (n = 257)	Hypertensive emergency (n = 356)
Headache				
Yes	266 (22.9)	94 (17.1)	96 (37.5)	76 (21.3)
No	161 (13.8)	70 (12.7)	37 (14.4)	54 (15.2)
Missing	736 (63.3)	386 (70.2)	124 (48.2)	226 (63.5)
Blurred vision				
Yes	69 (5.9)	18 (3.3)	29 (11.3)	22 (6.2)
No	181 (15.6)	62 (11.3)	75 (29.2)	44 (12.4)
Missing	913 (78.5)	470 (85.5)	153 (59.5)	290 (81.5)
Dizziness				
Yes	113 (9.7)	27 (4.9)	56 (21.8)	30 (8.4)
No	56 (4.8	19 (3.5)	24 (9.3)	13 (3.7)
Missing	994 (85.5)	504 (91.6)	177 (68.9)	313 (87.9)
Neurological deficit				
Yes	128 (11.0)	19 (3.5)	12 (4.7)	97 (27.2)
No	602 (51.8)	301 (54.7)	150 (58.4)	151 (42.4)
Missing	433 (37.2)	230 (41.8)	95 (37.0)	108 (30.3)
Altered consciousness				
Yes	191 (16.4)	85 (15.5)	15 (5.8)	91 (25.6)
No	624 (53.7)	285 (51.8)	158 (61.5)	181 (50.8)
Missing	348 (29.9)	180 (32.7)	84 (32.7)	84 (23.6)
Chest pain				
Yes	216 (18.6)	51 (9.3)	97 (37.7)	68 (19.1)
No	311 (26.7)	131 (23.8)	97 (37.7)	83 (23.3)
Missing	636 (54.7)	368 (66.9)	63 (24.5)	205 (57.6)
Palpitations				
Yes	71 (6.1)	14 (2.5)	45 (17.5)	12 (3.4)
No	244 (21.0)	92 (16.7)	84 (32.7)	68 (19.1)
Missing	848 (72.9)	444 (80.7)	128 (49.8)	276 (77.5)
Dyspnoea				
Yes	200 (17.2)	73 (13.3)	51 (19.8)	76 (21.3)
No	272 (23.4)	112 (20.4)	107 (41.6)	53 (14.9)
Missing	691 (59.4)	365 (66.4)	99 (38.5)	227 (63.8)

Values are numbers (percentages). Abbreviations: HE, hypertensive emergency; HU, hypertensive urgency.

Prospective study

During the inclusion period 59 patients met our inclusion criteria (*Figure 11.2*). Four patients spontaneously reported non-adherence. Of the remaining 55 patients, plasma drug levels were analysed. Based on drug levels, 14 out of 55 patients (25.5 %) were deemed non-adherent for at least one drug. Of these 14, seven patients were completely non-adherent for all measured drugs. Combined with the four patients who spontaneously reported non-adherence for all prescribed drugs, this means 30.5% of patients was non-adherent of which more than half (61%) fully non-adherent and 39% partially non-adherent. Of the 41 patients who gave informed consent for collection of clinical data, eleven (26.8%) were partially or totally non-adherent (*Table 11.4*), so there

Table 11.3: Additional investigations and results retrospective study

Additional investigations and findings	Total (n = 1,163)	No HE/HU (n = 550)	Hypertensive urgency (n = 257)	Hypertensive emergency (n = 356)
Laboratory investigation	1,053 (90.3)	454 (82.5%)	246 (95.7%)	353 (99.2%)
Cardiac markers	0 (7.6)	16 (2.9%)	13 (5.1%)	61 (17.2%)
Urinalysis performed	260 (22.4)	100 (18.2)	105 (40.9)	55 (15.4)
Proteinuria present	97 (8.3)	42 (7.6)	24 (9.3)	31 (8.7)
Chest radiography	424 (36.5)	185 (33.6)	99 (38.5)	140 (39.3)
Increased CTR	82	17	31	34
Ophthalmoscopy	116 (10)	11 (2.0)	77 (29.9)	28 (8.2)
Retinopathy	10			10
ECG	760 (65.3)	234 (42.5)	225 (87.5)	301 (84.6)
ECG abnormalities				
Left ventricular Hypertrophy	221	254 (9.9)	60 (23.3)	107 (30.1)
Signs of Ischemia	104	29 (5.3)	21 (8.2)	54 (15.2)
Echocardiography	90 (7.7)	14 (2.5)	23 (9.0)	53 (14.9)
Left ventricular hypertrophy	40	5	12	23
Head CT scan	304 (26.1)	44 (17.1)	26 (17.2)	234 (65.7)

Values are numbers (percentages). Abbreviations: HE, hypertensive emergency; HU, hypertensive urgency; CT-scan, Computed Tomography scan; CTR, cardiothoracic ratio; ECG, electrocardiogram.

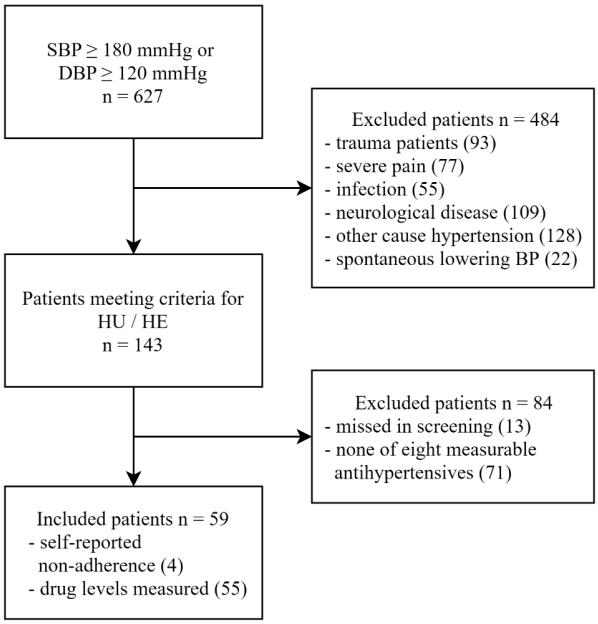


Figure 11.2: Flow chart prospective study. Abbreviations: SBP, systolic blood pressure; DBP, diastolic blood pressure; HU, hypertensive urgency; HE, hypertensive urgency

was no significant correlation between giving or withdrawing informed consent and adherence. Non-adherence was found in both HU (64%) and HE (36%). The difference in non-adherence between HU and HE was not statistically significant.

Non-adherence was the highest for spironolactone and amlodipine (*Figure 11.3*). No significant differences in clinical characteristics between the adherent and non-adherent group could be identified, aside from the higher number of previously prescribed antihypertensive drugs in the non-adherent group (3.7 vs. 2.7 antihypertensive drugs; p

Table 11.4: Patient characteristics including one year follow-up data prospective study

	Total (n = 41)	**Adherent** (n = 30)	**Non-adherent** (n = 11)
Male	19 (46.3)	14 (46.7)	5 (45. 5)
Mean (SD) age	65 (11)	66 (12)	62 (11)
Mean (SD) SBP (mmHg)	198 (16)	196 (17)	202 (17)
Mean (SD) DBP (mmHg)	106 (15)	104 (15)	111 (14)
Mean (SD) BMI	29.0 (5)	30.0 (5)	25.3 (4)
Diabetes	10 (24)	8 (27)	2 (18)
Previous cardiovascular disease	22 (54)	17 (57)	5 (46)
Chronic kidney disease	10 (24)	5 (17)	4 (46)
Non-adherence earlier reported	6 (15)	3 (10)	3 (27)
Hypertensive urgency Hypertensive emergency	20 (49) 12 (29)	13 (43) 8 (27)	7 (64) 4 (36)
Hospital admission	26 (63)	19 (63)	7 (64)
Change in drug regime after discharge (ED/ hospital)	19 (48)	14 (47)	5 (50)
Number (SD) antihypertensive drugs*	2.9 (1.4)	2.7 (1.0)	3.7 (2.0)
Taking ≥ 3 antihypertensive drugs	22 (54)	14 (47)	8 (73)
Prescribed antihypertensive drugs: ACE inhibitor Angiotensin II receptor blocker Calcium channel blocker Diuretic Betablocker	 20 (49) 13 (32) 20 (49) 21 (51) 25 (61)	 15 (50) 11 (37) 14 (47) 14 (47) 17 (57)	 5 (46) 2 (18) 6 (55) 7 (64) 8 (73)
Follow-up after one year Dead Alive, good blood pressure regulation Alive, poor blood pressure regulation Alive, blood pressure regulation unknown	 	 2 (7)[#] 19 (63) 6 (20) 3 (10)	 0 (0) 4 (36) 7 (64) 0 (0)

Values are numbers (percentages) unless stated otherwise. Abbreviations: SD, standard deviation; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; ACE, angiotensin converting enzyme; ED, emergency department. *Student t-test between adherent and non-adherent group differed significantly (p-value 0.04) [#]both unrelated to blood pressure (cancer).

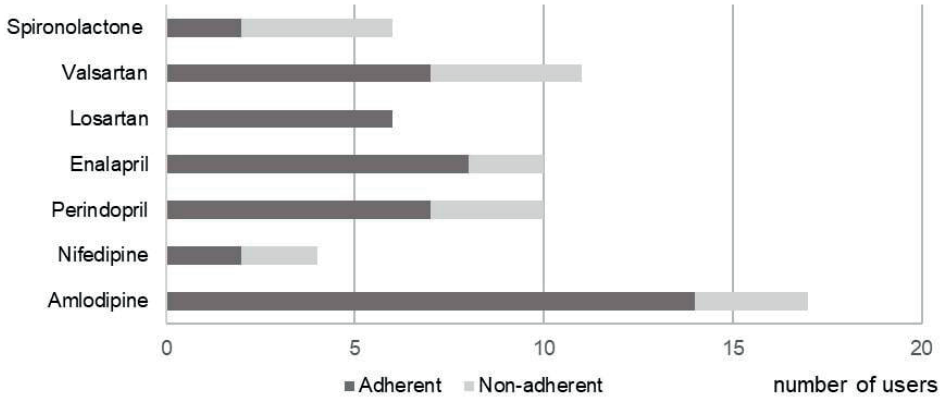


Figure 11.3: Adherence and non-adherence for each individual measured drug in the prospective study.

= 0.03). At discharge, the medication regime was changed in half of the patients. During subsequent visits to the outpatient clinic, two non-adherent patients had symptomatic hypotension, probably due to adherence. None of the patients visited the ED with a recurrent hypertensive crisis during the one year follow-up period.

DISCUSSION

In this study we showed that severely elevated BP is common in the ED and that the incidence is rising. Approximately half of the patients with severely elevated BP met the criteria of HU or HE, accounting for one in 200 ED visits. HE was more prevalent than HU. As expected, patients with HE were older and had more comorbidities (e.g. diabetes, hypertension, previous stroke). In only 5% of the patients suspicion of non-adherence was documented in the medical records, while in our prospective cohort we observed non-adherence in 30.5% of the patients. This discrepancy has important clinical implications, since an intervention improving adherence might be more beneficial than extending drug therapy by increasing doses or adding drugs. This latter strategy may potentially lead to severe side effects such as hypotension, as was seen in two of the patients in the prospective study.

In our study, the incidence of HU and HE together (0.5%) is in line with most other studies, where reported incidences range from 0.45% to 3.0%[65, 261, 265, 272]. The incidence of HE was higher than of HU, whereas most other investigators found a greater incidence of HU than HE[65, 265, 272, 273]. This probably relates to the fact that our hospital is a tertiary referral centre for specialised treatment such as thrombectomy for ischemic stroke[274], craniotomy for intracerebral haemorrhage or thoracic surgery for aneurysms. Only one

other study on hypertensive crisis in the ED found more HE then HU with comparable patient/hospital characteristics[261].

We noted an increase in incidence of hypertension-related visits to the ED of 7.7% per year, which is probably due to the increase in hypertension in the general population[275]. This percentage is comparable to the increase reported by *McNaughton et al.* in 2015, who reported an increase of hypertension-related ED visits of 5% per year in the period 2006-2012[262].

Patients with HE presented most often with ischemic and haemorrhagic stroke, acute pulmonary oedema and myocardial infarction. Their clinical symptoms mostly fitted the subtype of HE. Headache was the most common symptom and was, in most cases, a sign of HU and not of HE in line with earlier studies[265, 276]. Approximately five percent of the patients with severe hypertension had no clinical symptoms of organ damage, but were diagnosed with either HU (29%) or HE (7%) after extensive testing. This implies that assessment of symptoms alone is insufficient to rule out HU and HE. Previous studies reported higher proportions of patients without symptoms, which may be partly explained by the BP criteria used[260]. Comparing HE and HU, we found that HE patients were older, had higher SBP and DBP, and more often smoked, which is in line with earlier studies[260, 266, 267, 272]. Approximately two third of patients with HE or HU had a previous history of hypertension, which is relatively low compared to percentages found in earlier studies (70% to 90%)[260, 265-267]. Despite the differences between HU and HE, no (combination of) factors could be identified that distinguishes HE from HU without the use of additional testing. Therefore, the standard workup, which includes blood tests and ophthalmoscopy, should always be followed in order to differentiate between HE and HU, if clinical signs do not yield the diagnosis HE. In our hospital, the guideline was not consistently followed. Especially ophthalmoscopy was performed in a low percentage of cases. It should be kept in mind that in HE it might not have been necessary to confirm the diagnosis when another kind of organ damage was identified, but in HU it should have been 100% to rule out HE. This low percentage of following the guidelines is in line with an earlier study that found even lower numbers: in only six percent of all patients presenting with severe hypertension all tests were performed as recommended in the local guideline[259]. Standard workup should include repeated measurements of blood pressure since blood pressure often lowers spontaneously and/or in response to pain or stress relief[277]. Although repeated measurements were performed in most instances, enabling us to select patients with spontaneous blood pressure reduction as a separate category "non-HU non-HE", follow-up measurements were not always reported in the electronic patient record. Since carrying out this study, the American guideline has been altered to consider all severe hypertension not HE as "markedly elevated blood pres-

sure", thus abandoning the terminology of HU. In this study, we employed the existing categorization according to the guidelines valid at the time[60, 263, 264].

Many studies reporting about the incidence of HU and HE suggest that non-adherence is a contributing factor to the development of HU or HE, but studies directly assessing this possibility are limited. In a recent study, drug levels of all patients on antihypertensive drugs visiting the ED (for any reason) showed 28% non-adherence (undetectable drug levels), and non-adherence was associated with higher BP levels[278]. The assay used in this study, however, was validated using clinical samples from hospitalised patients obtained shortly after drug intake and as a consequence drug levels measured at time points more than 12 hours after intake could have been false negative, overestimating the prevalence of non-adherence[279].

In a study including patients with stroke, the odds ratio for developing a stroke as a result of non-adherence to antihypertensive drugs ranged from 1.7 to 2.7, depending on the number of years antihypertensive drugs had been prescribed[265, 269]. A small prospective study defined non-adherence as a risk factor for the development of hypertensive crises, where non-adherence was defined solely on reporting by patients and physicians[280].

Our study is the first study focusing on patients with severe hypertension in the ED using a well-validated LC-MS/MS method, which is the most reliable method to asses adherence[281]. We found that 30.5% of the patients were indisputably non-adherent, when also taking into account the patients who self-reported non-adherence at inclusion. This rate is in line with earlier studies performed in uncomplicated hypertension and resistant hypertension[282-288]. Therefore, our findings imply that improving adherence is of major importance, especially since suspected non-adherence was documented in the electronic medical record in only five percent of patients in the retrospective study. Two out of nine patients who received follow-up in the outpatient clinic developed hypotension after the ED visit. These complications, presumably caused by using all previously and newly prescribed drugs, might be prevented by adequate assessment of adherence in the ED. Two observational studies indicate an improved BP control when providing feedback on undetectable drug levels[289, 290]. Although this was not studied in a controlled way, it implies that immediate measurement and feedback at the ED might be an efficacious approach to improve adherence and consequently BP control.

When investigating the differences between the adherent and non-adherent patients, we found that non-adherent patients had been prescribed more antihypertensive drugs than adherent patients. Studies have shown that an increase in the number of prescribed drugs results in more non-adherence[291, 292]. The drug for which most patients

were non-adherent was spironolactone. Spironolactone is commonly used in resistant hypertension after the PATHWAY-2 trial[293]. Our findings suggest that the reason the BP target in these patients was not previously reached is non-adherence, whereas not reaching the BP target urges the physician to prescribe spironolactone.

Strengths and weaknesses

This study has strengths and limitations. The first strength of our retrospective study is the large number of patients for whom most of the relevant parameters were known. Secondly, the results from our retrospective analyses can easily be generalised to other countries, since the Dutch population is comparable with populations of most western countries, including the USA in terms of ethnicity, lifestyle, habits and disease incidence and prevalence. Our study population comprised of multiple ethnicities with patients born in more than 41 countries.

Our study also has limitations. First, in the retrospective study we included only patients with both high SBP and DBP using cut-off points of ≥ 180 mmHg and DBP ≥ 110 mmHg, respectively. With this approach we potentially missed patients with HU or HE with an isolated high SBP or DBP, resulting in an underestimation of the incidence. However, by applying both SBP and DBP we limited the number of patients with increased BP due to pain or stress. In the prospective study we chose threshold values of SBP ≥ 180 mmHg or DBP ≥ 120 mmHg, limiting direct comparison of the retrospective and prospective study. However, these threshold values did not influence the percentage of non-adherence. Also, there are no clearly defined cut-off values at which the risk of acute end-organ damage is absent or present. The acceleration of BP rise is more important than absolute BP values.

Secondly, due to the nature of care in the emergency setting, we encountered missing data. This may have led to underestimation of suspicion of non-adherence, although spontaneous report is more likely in presence than in absence of suspicion of non-adherence. However, lack of spontaneous reporting of suspected drug (non-) adherence also gives important information on physicians' awareness of this problem. In addition, even when physicians are at least considering the chance of non-adherence as underlying cause, they tend to overestimate their patients' drug adherence[294]. Therefore, we think the retrospective study is representative for how the issue of drug adherence is handled in daily clinical practice. Finally, the study was executed in an urban hospital that also functions as a tertiary care centre. In that capacity, a proportion of the patients had a complex medical history with multiple comorbidities and may therefore not be generalizable to secondary care centres, as also reflected by the higher incidence of HE than of HU.

Considering the prospective study, we were the first to prospectively investigate drug non-adherence in patients with a severely elevated blood pressure at the ED using a well-validated LC-MS/MS. By ensuring that drug levels were analysed independent of obtaining informed consent, we were able to avoid a potential selection bias. The number of patients not giving and especially withdrawing consent was surprisingly low. However, this study also has several limitations. The study was underpowered to compare characteristics of adherent and non-adherent patients, since the proportion of non-adherence was lower than expected. Using our assay, we could measure seven antihypertensive drugs. We chose this specific selection of most commonly prescribed antihypertensive drugs, because to use plasma drug levels in clinical practice the assay needs to be extensively validated. Diuretics other than spironolactone were not measured as they are not detectable for 24 hours after intake allowing measurement at a random time point (the ED visit). By not measuring diuretics and beta-blockers, drugs known to be associated with considerable non-adherence due to side effects, non-adherence might have been under-estimated[295]. However, all measured drugs are in the top 100 of most used drugs in The Netherlands and all of the chosen ACE inhibitors and angiotensin II receptor blockers are available in fixed-dose combinations. Being non-adherent for one of the measured drugs implies that adherence to unmeasured drugs is also questionable. In addition, since both the nominator and the denominator of the calculation of non-adherence depend on the choice of antihypertensive drugs, this limitation did not lead to a bias. A final issue to discuss is white coat adherence: in theory, patients could have taken their antihypertensive drugs just before going to the ED, but this might be more common during regular visits to the clinic than to the ED.

CONCLUSION

Severely elevated BP is a common health problem and the incidence is increasing, resulting in frequent ED visits and high economic burden. We showed that all patients visiting the ED for suspicion of HE should receive a full workup, regardless of their clinical symptoms. We found in the prospective study that 3 in 10 patients were non-adherent for antihypertensive drugs, while in the retrospective study only 1 in 20 patients physicians actively recorded non-adherence as a potential cause. Distinguishing between non-adherent and adherent patients is crucial, since treatment strategies differ. Therefore, ideally a point-of-care-test would be developed to enable direct assessment of adherence in order to adjust the treatment strategy accordingly.

Section IV

Summary


chapter
12

General Discussion, Future Perspectives and Conclusions

In the last years, the emergency department (ED) landscape has changed dramatically. The number of patients presenting to the ED decreased slightly, however, the number of older and more severely ill patients increased[2]. Also, the number of patients with chronic diseases is increasing. Of the general population, 58% has at least one chronic condition. In patients of 75 years and older, 70% has at least three[296]. The combination of an older ED population with severe illness and multiple comorbidities makes medical care in the ED increasingly complex. This extends the time needed to fully evaluate and treat a patient in the ED, resulting in a prolonged length of stay (LOS), which is an important factor for crowding[297, 298]. The recent emergence of acute medicine as a subspecialty of internal medicine has led to more expertise, more decision-making power and competence in the ED. Yet, this does not fully compensate for the still increasing complexity of ED medical care[7]. As a consequence, it remains difficult to predict morbidity and mortality in the ED, and research is badly needed to meet this challenge.

The research presented in this thesis aims to contribute to the field of acute medicine by developing, evaluating, validating and improving prediction models for admission, morbidity and mortality in the ED. Another aim is to provide tools to recognize deterioration of ED patients in a timely manner, and to explain why in some groups of patients deterioration is hard to recognise. The studies were performed in non-trauma patients admitted to an ED, in both the general population and in specific patient groups (e.g. elderly, patients with infections).

This discussion addresses the clinical implications of the findings, methodological aspects of the prediction models, discusses limitations of the studies, provides recommendations for future research and ends with an overall conclusion. The following specific questions were addressed and will be discussed:

1. Can we reliably predict hospital admission in patients in ED patients?
2. Can we reliably predict mortality in patients in the ED based on readily available predictors?
3. Can we use prediction tools to predict outcomes in specific patient populations?
4. Does non-adherence to guidelines and treatment have negative effects on patient outcomes?

Overview of the main findings per research question

Can we reliably predict hospital admission in ED patients?

Primarily, we performed a systematic review to identify current admission tools. We used the gained knowledge from the systematic review to develop and validate the CLEARED tool for admission in an older population at the emergency department. This tool had a good performance and held up through internal and external validation. The CLEARED tool is potentially capable to guide swift admission.

Can we reliably predict mortality in patients in the ED based on readily available predictors?

We evaluated existing prediction models for short-term mortality, developed for use in the ED, in a systematic review. Almost all prediction tools we evaluated in this systematic review had a risk of bias in the developmental stage. Furthermore, several models did not use readily available parameters. None of the prediction model performed satisfactorily in predicting mortality. Subsequently, we externally validated a model that used six laboratory parameters and the patients' age to predict mortality which showed a good discriminative performance. Afterwards we adjusted this model by only including laboratory parameters that could be tested with POCT. This yielded comparable performance, while the model consists parameters that are readily available at presentation at the ED .

Can we use prediction tools to predict outcomes in specific patient populations?

We tested whether current early warning scores can also be used in specific patient populations. In older patients current EWS were mediocre at predicting 30-day mortality. NEWS however was the best prediction tool. The addition of age and the APOP screener as a proxy for frailty significantly improved the predictive performance of EWS in this population.

In patients suspected of sepsis, the predictive performance of NEWS was compared to qSOFA and SIRS in predicting short term mortality. The NEWS was found to be a better prediction model than these sepsis specific models. In these patients we found that normothermia was associated with a higher mortality. This was confirmed in a subgroup of patients with bloodstream infections.

Does non-adherence to guidelines and treatment have negative effects on patient outcomes?

We showed that adherence to antibiotic guidelines in the ED was low. Both over- and undertreatment occurred regularly. However, neither of them had an effect on mortality rates. Antibiotic prescription probably depends on disease severity, because overtreatment was more often observed in patients who were more severely ill and undertreatment was more observed in less ill patients.

Non-adherence to antihypertensive treatment is a major contributing factor to the occurrence of hypertensive crisis. However, it is regularly not recognised by physicians. Awareness of this phenomenon could lead to prevention of overtreatment of hypertension at the ED.

CLINICAL IMPLICATIONS

Admission

In the **first section of this thesis**, we aimed to answer the question whether we can reliably predict hospital admission in ED patients based on readily available predictors. To accomplish this we first conducted a systematic review in **Chapter 2** to find existing prediction tools for admission. In our review, we found that of presently published tools, the Glasgow admission prediction score (GAPS) could predict admission best[81]. While evaluating existing prediction models of admission, we found that they frequently used parameters that are not readily available, such as whether the patient has memory problems or difficulty walking, or uses more than six different types of medication. Also, these models were often not validated or calibrated, making them less generalizable to other hospitals or treatment settings. Only four studies in our review were performed in an older patient population, and one study compared the adult to the elderly population[35, 82, 83, 87, 89]. We therefore also focussed research on the elderly.

Elderly patients in the ED

Care for older patients in the ED is challenging. Their needs differ from the general adult population and their risk for an adverse outcome (e.g. mortality) is higher[72]. Patient assessment can be laborious and results can be incomplete. Medical history can be hampered as a result of reduced cognitive performance or delirium, requiring collateral history[299]. In **Chapter 6** we showed that vital signs have a different predictive value on 30-day mortality in older patients than in the general population, making them harder to interpret. This is a result of, amongst others, the presence of co-morbidities and polypharmacy. Therefore, this population will likely benefit from a prediction tool that takes these aspects into account.

Prediction of admission in elderly patients

Taking the findings of **Chapter 2** in account, and considering clinical applicability, we aimed to provide a methodologically sound prediction tool. We constructed a prediction tool for admission in **Chapter 3**, called the "CalcuLation of the Elderly Admission Risk at the ED" (CLEARED)-tool. This prediction tool was mainly based on vital signs. CLEARED has a high predictive value for admission for patients aged 70 years and over in presenting to the ED, and was validated and calibrated in two other hospitals, showing good performance. As it only contains easily obtainable parameters, it can give predictions within minutes after arrival at the ED. Using CLEARED may accelerate the ED process by shortening the time to disposition decision, leading to shortened LOS. An early disposition decision is beneficial as a prolonged LOS in the ED can lead to profound negative outcomes in older patients, such as increased susceptibility to delirium, an increase in

functional dependency and mortality[23, 300, 301]. When a patient is identified as having a high risk of admission, the admission process can already be started. Patients can either be admitted prematurely to an acute medical unit (AMU) or to the inhospital ward awaiting further evaluation and treatment, or the patient remains in the ED until full evaluation has been completed, whilst the admission is being arranged. Older patients are a relevant and interesting target for interventions such as the CLEARED, because this group has a high chance of negative outcomes after an ED visit, and admittance numbers and re-visit rates are high[302-304].

Mortality

Older patients have, in comparison to younger patients, a higher mortality risk. In Erasmus MC overall 30-day mortality is 2.6% following an ED visit. The 30-day mortality rate is six times higher in patients over 70 years of age, as compared to younger patients (10.9% vs. 1.8%). This is a result of a combination of age, presence of comorbidities, loss of resilience and an atypical disease course[305, 306]. While it is important to predict admission, it is of greater importance for patients and their relatives to predict the chance of dying. Quick identification of the most severely ill patients can help focus treatment resources, and therefore prevent loss of life. In **the second part of this thesis** we looked into the question whether mortality can be reliably predicted in ED patients based on readily available predictors. We therefore conducted a systematic review of prediction tools for short-term mortality in **Chapter 4**. We found 22 models predicting either 24-hour, 5- 10- and 30-day and inhospital mortality, and the full model described by *Brabrand et al.,* which uses vital signs, age and loss of independence as parameters, performed best[131]. Regrettably, to date none of these models have been implemented in clinical practice.

Some of these models use laboratory data. This improves accuracy, but makes the process more time-consuming, while in the ED it is essential that prediction models are easily applicable and use readily available parameters. In **Chapter 7**, we evaluated an existing prediction model, developed by *Asadollahi et al.*, based on six different laboratory tests and the patients' age[195], which showed good performance. Although conventional laboratory test results are not readily available, test results using point-of-care tests (POCT) are. Therefore, the model was adjusted by only including parameters which could be tested with POCT. This new model yielded comparable performance, making it more suitable for the ED setting. POCT as an alternative for regular laboratory results might improve accuracy of prediction models with only limited extend in time. Its place still needs to be decided, as it can either be used in specific patient groups, or in all patient groups that require laboratory testing.

Predicting mortality using Early Warning Scores

Early Warning Scores (EWS) are increasingly used in the ED, even though EWS were initially developed for identifying possible deterioration of hospitalised patients based on inhospital data. EWS can not only identify direct risk on deterioration, but also prognosticate long-term outcome[307]. Presently, several EWS are being used[308, 309]. With the introduction of "The Third International Consensus Definitions for Sepsis", the quick Sepsis related Organ Failure Assessment (qSOFA) was introduced[55]. This is a model that predicts mortality in patients with infections. qSOFA was also developed and validated in hospitalised patients, whereas most patients with sepsis present to the ED. Therefore, in **Chapter 5**, we investigated the performance of the National Early Warning Score (NEWS), Systemic Inflammatory Response Syndrome (SIRS), and qSOFA in predicting 10- and 30-day mortality in patients suspected of having an infection in the ED. NEWS had the best discriminative value, whereas SIRS and qSOFA either had high sensitivity with low specificity, or vice versa. Moreover, NEWS was the only tool that was adequately calibrated in contrast to the sepsis-specific tools. We also confirmed that the best cut-off point for deterioration using NEWS was ≥ 7 as was stated by the United Kingdom National Health Service (NHS)[158]. After publication, our findings have been confirmed by several other studies[177, 310, 311]. In an effort to increase the validity of a sepsis specific score for deterioration, one study adjusted the qSOFA by the addition of oxygen saturation[312]. Another study added lactate to the qSOFA[313]. Both of these adjustments increased the predictive performance of the qSOFA for deterioration, yet were still inferior to the NEWS in our study. According to these results, there is no advantage in using qSOFA over NEWS, even if the qSOFA is specifically developed for use in sepsis. Other studies have shown that early treatment in sepsis is paramount. So, if NEWS were used as prediction tool in sepsis, it could be used, not only to prognosticate a disadvantageous course of disease, but also to trigger life-saving interventions as suggested by the Surviving Sepsis Campaign (e.g. blood cultures, early administration of antibiotics)[314].

To identify patients at the verge of deterioration, EWS mainly use vital signs such as body temperature. The presence of fever is historically one of the best known markers of infections. Historically, fever was considered a disease instead of a symptom. It was not until between the 17th and the 18th century fever was considered a symptom associated with internal modifications of the body[315]. An elevated body temperature (i.e. fever) is almost always a sign of inflammation and is most often caused by a bacterial or viral infection[316]. In EWS, a higher score for risk of deterioration is assigned to a deviant body temperature. In **Chapter 9** we studied the association between body temperature, infection and outcome. While it is counter-intuitive, patients in the ED without a fever who had an infection had higher mortality rates, which is not fully explained. It is hypothesised that the severity of their illness is underestimated due to the absence of

a fever, leading to a delay in treatment. In older patients, infections are frequently not accompanied by fever[317-319]. Patients who are treated with immunosuppressants (e.g. patients with organ transplants, patients who undergo chemotherapy) are instructed to take their temperature and present to the ED when having a fever[320, 321]. Nevertheless, as the height of the body temperature varies throughout the course of a disease, the patient may subsequently present with a normal body temperature[322]. However, in these patients the course of the disease is more severe. Another potential explanation is that antipyretics reduce fever[323]. As fever is accompanied by an increase in heart rate and respiratory rate, absence of fever may mask severity of disease in both clinical judgment and EWS, and could therefore be a possible confounder. Further research is needed to investigate which of these explanations contributes most to these findings in order to improve the outcome of patients that have an infection, but no fever. The use of the historical highest body temperature of the patient (i.e. the body temperature measured before presenting to the ED) in EWS, should be investigated.

Screening the elderly in the ED

An increasing share of patients in the ED are older patients. Most prediction models used to detect deterioration in clinical practice have not been validated in an older patient population. It is important to realize that normal values of vital signs are age-dependent and subgroup-specific (e.g. children, pregnant women). Not only the vital signs differ, but also the pathophysiological reactions to disease differ in these subgroups. This has led to adjusted EWS such as the PEWS for pediatric patients and the MOEWS for obstetric patients[45, 47]. However, for older patients no such adjustments have yet been made based on age specific vital signs so far. Although some early warning scores, e.g. the Rapid Emergency Medicine Score (REMS), attribute mortality risk dependent on age[143]. In **Chapter 6** we evaluated the performance of four EWS (i.e. NEWS, MEWS, REMS, RAPS) predicting mortality in a population aged 70 years and over. NEWS performed best in predicting mortality in these elderly patients, albeit still suboptimal. Mortality rates were higher than in a general population, and older patients who were considered frail had an even higher risk for 30-day mortality. In older patients vital signs may seem less alarming, resulting in underestimation of the severity of illness and subsequently the risk of deterioration. While normally patients suffering from sepsis present with fever, the older patients often display an altered consciousness or delirium as pivotal symptom of sepsis. Vital signs may also be altered by the use of specific medication. For example, beta-blockers and calcium channel blockers can prevent tachycardia in shock and NSAIDs can suppress the development of fever in sepsis. This often results in impaired recognition of severely ill older patients, with potential detrimental consequences. In an attempt to improve the predictive performance of EWS, we successfully added age and a proxy for frailty. The proxy for frailty we used was the APOP screener, which was

developed in the Leiden University Medical Center and Alrijne hospital and validated in Bronovo and Erasmus MC[184]. The APOP screener is currently implemented in more than 50% of the hospitals in the Netherlands[324], where it is used to identify frailty with the objective to start interventions to reduce functional decline in older ED patients. Our data suggest that frailty is also a major determining factor for 30-day mortality.

An important part of the assessment of frailty is mobility. Possibly, the assessment of pre-morbid mobility is also of prognostic value in the general population and maybe mobility should be considered as a vital sign[325-327]. A convenient way to assess mobility is by performing the "Timed Up and Go Test". This test is performed by measuring the time it takes a patient to get up from a chair and walk three meters without encouragement or physical aid[328]. However, assessment of mobility in this setting also depends on other vital signs, like blood pressure and heart rate.

From clinical perspective there is an urgent need for decent prediction tools to identify older patients at risk of adverse outcomes. The objective is to improve the outcome in this population. From a policy-making point of view this urge is also felt. In 2008, the Dutch Safety Management Program (VMS) was announced. This program introduced a screener to identify older hospitalised patients at risk of adverse outcomes, such as functional decline, risk of delirium and mortality. This VMS screener studies four domains (i.e. delirium, fall risk, physical decline and malnourishment) and is a reliable instrument when used in the wards to detect frail older patients[329]. However, according to a recently published study in 249 older patients it has a very low positive predictive value in the ED, making this screener unsuited to identify patients at high risk for adverse outcomes in the ED[330]. This study also demonstrates the frequently found phenomenon that the performance of a prediction model is dependent on the setting in which it is used.

Predictors of medication prescription and non-adherence

The treatment of patients often involves prescription of some sort of medication. While the aim of this treatment is health benefit, medication errors can cause considerable damage, which forms an important public health hazard. The European Medicines Agency defines a medication error as an "unintended failure in the drug treatment process that leads to, or has the potential to lead to harm to the patient"[331]. The most common - preventable - errors are mistakes in the prescribing, dispensing, storing, preparation and administration of medication. Some consider non-compliance of a patient as a medication error as well[332]. A substantial number of ED visits are in some way related to or caused by medication errors happening at home[333]. A notable part of these ED visits and admissions are preventable[334]. Medication errors do not occur only at home. Also inhospital medication errors form a big problem with potentially fatal consequences.

Patients in the ED are also at risk for medication errors[335]. Sometimes home medication usage and allergies are not known at presentation at the ED. In acute situations not always all precautions regarding medications can be followed. Frequently, medication orders are initially done verbally instead of in writing. Also, protocols and guidelines are not always followed, either due to lack of knowledge or due to bias depending on severity of disease. In a hectic situation such as the ED, the risk of these kind of errors is increased. In **the third section of this thesis** we describe some factors of influence on drug prescription, adherence to guidelines, and adherence to treatment by the patient.

Adherence

In **Chapter 11** we studied non-adherence to antihypertensive therapy in the ED. We found that non-adherence to antihypertensive agents is an important factor in majority of patients visiting the ED with hypertensive urgency or emergency. There were discrepancies between the proportion of patients of which doctors contributed the hypertensive crisis to non-adherence, and the actual number of patients which were non-adherent to antihypertensive treatment according to medication levels in blood plasma. Currently, it is possible to measure non-adherence using a dried blood spot method by a single finger prick. Its use in the ED should be further expedited. When these tests are available as POCT, the result will be available within minutes. In case of non-adherence re-starting medication is most likely sufficient, but it is of importance to determine the reason why medication was not used. In case the patient is adherent to treatment and the blood pressure remains high, adding a different kind of medication should be considered. Our study shows that physicians should consider non-adherence more often.

Factors affecting drug prescription

Antibiotics are frequently prescribed in the ED and form the cornerstone in the treatment of infections, which are when culminating into sepsis, potential life-threatening conditions[336]. The mortality rate of sepsis is high, and numerous studies have shown that early treatment with antibiotics is paramount. In order to reduce mortality of sepsis, the Surviving Sepsis Campaign was introduced. The resuscitation bundle includes administration of antibiotics within one hour (i.e. the golden hour)[314], as a landmark study showed that delay of adequate treatment increased mortality by 7% per hour in patients with septic shock[337]. However, subsequent studies were never able to reproduce this effect, and, a study on prehospital admission of antibiotics in patients with sepsis did not show a lower mortality[243]. This implies that there are patients who benefit from early administration of broad-spectrum antibiotics, and that there are also patients in whom it is justified to await results from diagnostics to narrow antibiotics to a more tailored

therapy. However, currently this decision is mostly based on the clinician's judgement of the patient.

In order to improve understanding of the process leading to the prescription of antibiotics, we investigated factors that influence a clinician's decision for antibiotic treatment. **Chapter 9** shows that the presence of fever is associated with the prescription of antibiotics. As described earlier in this discussion the absence of fever was associated with higher mortality rates in patients with infections in the ED. One hypothesis is that in the absence of fever physicians are less inclined to start antibiotic treatment, as they possibly underestimated the severity of illness in patients. However, the outcome in these patients was worse, and undertreatment might play an important role. We recommend that when a healthcare provider suspects a severe infection and therefore performs blood cultures, the presence of normothermia should not preclude the use of antibiotics.

The appropriateness of empiric treatment

To guide antibiotic therapy blood cultures are taken to determine the species of bacteria and their resistance, but cultures require time to become positive for growth. As patients suffering from sepsis need immediate treatment and blood cultures cannot be awaited, empirical antibiotic treatment is started based on the suspected origin of the infection - if present - and the suspected causing species of bacteria[338]. This 'antibiotic stewardship' aims to contain the development of antimicrobial resistance and provide guidelines for optimal use of antibiotics, taking resistance surveillance data into account[339]. These resistance data can either be local or national. In hindsight this empirical treatment can be classified as appropriate if the bacteria is susceptible to the chosen antibiotic, or inappropriate if the bacteria are resistant[59].

In **Chapter 10,** we compared the initial antibiotics with the results of blood cultures that later became positive, and found that over 70% patients in the ED were prescribed appropriate antibiotics. Unexpectedly, appropriate or inappropriate treatment had no effect on mortality. This might have been due to the heterogeneity of our cohort. Appropriately treated patients were more severely ill than inappropriately treated patients. In spite of correction for multiple confounders, residual confounding remained an issue. However, when we performed a subgroup analysis in a more homogenous group of patients treated with only one type of antibiotic, a relation between inappropriate treatment and mortality was confirmed as expected. This study showed that retrospective research in heterogeneous groups is very difficult. Many confounders are of influence on the parameters investigated and are frequently hard to identify. Therefore, we recommend that research on treatment effects in observational data can be done best in well-defined homogenous patient groups. A recently published review and meta-

analysis confirmed the association between appropriate antibiotics and a favourable outcome in patients with a bloodstream infection. This review reported on studies with more homogeneous groups than in our cohort, namely, in patients with severe bacterial infections[340]. Appropriately treated patients had more favourable outcomes, highlighting the importance of appropriate therapy in patients with a severe infection, which can be achieved by adherence to guidelines.

(Non-)adherence to treatment guidelines

Once the origin of an infection is established and empirical treatment is deemed necessary, the choice of the type of antibiotic agent and its dose is based on regional or national guidelines. In **Chapter 8** we investigated adherence to antibiotic guidelines. To our surprise, physicians in the ED were predominantly non-adherent to the guidelines when initiating antibiotic treatment. The type of antibiotics was more frequently divergent from the guideline than the dose. This non-adherence resulted in both undertreatment and overtreatment. Undertreatment in the ED was mostly the result of not initiating antibiotics. Remarkably, non-adherence had no effect on mortality. Patients who are undertreated are less ill, and have a lower chance of dying. However, undertreatment results in a higher chance of inappropriate therapy and, as we have shown in **Chapter 10**, a higher chance of negative outcomes. Overtreatment was associated with a higher disease severity. However, overtreatment had no advantage to the patient and thus should be avoided. Guidelines aim to achieve adequate coverage. In illnesses in which different bacteria cause different severity of illness, such as in pneumonia, antibiotic advice differs in accordance to the severity of illness, and scoring systems are used (e.g. CURB-65)[341]. However, in most cases the severity of illness is not a result of the bacteria, but of host factors. Adherence to these guidelines will therefore likely result in adequate treatment, making overtreatment not additionally beneficial, since adequacy is already met.

Improving prescription of antibiotics

The key to improve sepsis treatment is not increasing the dose or combining antibiotics, as this would only increase the risk of side-effects and resistance. To improve sepsis treatment, appropriate antibiotics should be prescribed earlier based on a test that ideally identifies the specimen of bacteria including their antibiotic susceptibility, to allow targeted therapy. Presently blood cultures are the gold standard to confirm bloodstream infections. The time from inoculation to blood culture positivity depends on the growth rate of the bacteria and the initial load of the pathogen. Unfortunately, this process takes at least a day, but it is frequently more time-consuming[342]. It would be beneficial to determine the causative bacteria together with its sensitivity to antibiotics as early as possible. Currently a number of molecular microbial essays and real-time PCR are avail-

able to faster diagnose bloodstream infections using peripheral blood samples. This can substantially reduce the need for or duration of empirical treatment. An example is the Sepsis@Quick essay which is faster in identifying bacteria than blood cultures, and also showed to have a positive effect on sepsis-related mortality[343]. However, due to several constraints these molecular techniques are not regularly used in the microbiological laboratory[344, 345].

Antibiotic treatment should only be started if a disease is the result of a bacterial infection. Several biomarkers can discriminate between bacterial and nonbacterial disease. The hiTEMP study aimed to investigate whether the use of such biomarkers, such as procalcitonin, TRAIL and IP-10 could correctly discriminate between bacterial and nonbacterial disease. The investigators hypothesised that an elevated procalcitonin level could guide the initiation of antibiotic therapy. However, this approach failed as it did not result in any difference in antibiotic prescription or mortality[346]. This is probably a result of several limitations. A selection bias was introduced as they included only patients with fever instead of all patients suspected for infection. Furthermore, withholding antibiotic treatment while waiting for the result of the laboratory tests is not feasible when the patient is suspected for life-threatening infections. This shows that there is still need of alternative methods to identify patients in the ED in need of antibiotic therapy. Speeding up the process of identification of pathogenic bacteria and establishing their antibiotic sensibility may result in a substantially earlier start of targeted antibiotic treatment, reducing mortality and other adverse outcomes.

PREDICTION RESEARCH IN ACUTE MEDICINE

Prediction models

Prediction models are increasingly used in clinical practice. Preferably, these models aid clinical decision-making and improve effectiveness and quality of care. However, we found that many prediction models developed for use in the ED were of mediocre quality. We provide insight in what these prediction models lack, and how these flaws should be addressed.

Model development

Model development starts with defining the research question and the outcome of interest. Subsequently, candidate predictors are selected. Selection of candidate predictors can be based on the used dataset, but in general selection based on literature or expert opinion is preferred[347]. There are several techniques to design a model from a set of candidate predictors. For instance, this can be done by stepwise model develop-

ment or by univariable regression analysis. In stepwise model development, backward elimination is a frequently used strategy in which all potential predictors are included in the model and the least predictive parameters are excluded stepwise. The selection stops when otherwise the deletion of more parameters leads to a significant loss of fit. However, stepwise selection may lead to biased regression coefficients and overfitting[148]. Overfitting is defined as "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably"[348]. With the use of univariable analysis the most potent predictors can be identified, which then are included in multivariable analysis. Nevertheless, when large databases are used it is likely that most candidate predictors have a significant relationship to the outcome. To account for this, better solutions have been proposed, such as the Akaike information criterion (AIC), which is an estimator of the relative importance of the predictor while preventing overfitting[349]. As we also possessed a large database, we used the AIC to select parameters for our CLEARED tool. Based on the systematic reviews in this thesis, no other study applied this estimator to develop prediction models for admission or mortality for use in the ED. We recommend that this modelling strategy should be used in developing future prediction models.

Another important, but frequently ignored factor in developing prediction models is the number of events per variable. The number of predictors (variables) that can be considered for model development is limited by how frequent the given outcome occurs in the dataset (events). In general, the number of events per variable should be at least ten, but larger numbers are preferred[149]. A lower number of events per variable may lead to misleading associations and less accurate regression coefficients. In many studies the number of events per variable was not considered, or was not mentioned. For CLEARED, 13 variables were tested and 4,079 patients were admitted. Therefore, the number of events per variable was 313 which is far above the recommended minimum of 10 events per variable. In **Chapter 5** in the model predicting mortality this number was 70, as there were 490 deaths and the number of included parameters in the prediction model was seven. Thus, it is unlikely that these models are overfitted to the development data.

Finally, the presence of missing data is of influence on model development. As model development is most frequently performed in existing cohort studies, missing data are common. There are several options to deal with missing values. Often a complete case analysis is performed, which means that every patient with missing values is excluded. However, this may substantially reduce the study size and may even result in a drop of events per variable below ten. Subsequently, this may lead to biased estimates, specifically when data are missing not at random. There are several other options to deal with missing values. One possible approach is replacing the missing values by the mean or

median of the variable. Other options are single or multiple imputation techniques[350]. Multiple imputation is considered the best way as it generates multiple complete datasets, taking into account that predictors are correlated to each other and can vary among patients and the uncertainty in the imputation model[351]. We opted to use this technique in many of the studies in this thesis to prevent bias.

Model performance

Once a model is developed, its performance should be evaluated in independent data. Performance is usually described in terms of discrimination, i.e. the ability of the model to distinguish the high-risk from the low-risk population. We found that in current prediction tools for use in the ED, when predicting admission, discrimination ranged from 0.63 to 0.88 and from 0.63 to 0.93 when predicting short-term mortality. However, discrimination measures in model development studies are often too optimistic as the regression coefficients of the predictors reflect the underlying population. This leads to overfitting. Subsequently, it is important to consider the clinical applicability of the model as well as the clinical usefulness: will the use of the model lead to better decisions in clinical practice?

Validation

To assess model performance beyond the development data and assess potential overfitting, a prediction model needs validation. This can be internal validation and external validation. Internal validation aims to provide more accurate estimates of model performance in new patients from the same setting. During internal validation the model is tested in (a sample from) the original study population. This can for instance be done by split-sample techniques, cross-validation or bootstrapping. The recommended method is bootstrap resampling, in which the model is tested in randomly sampled datasets from the original dataset[114]. Internal validation is a recommended step before external validation. External validation should be carried out to test whether the model is generalizable to other hospitals or other settings. During external validation both the discrimination and the calibration should be assessed, of which the latter is rarely performed[352]. Validation is frequently lacking in prediction models developed for use in the ED. Additionally, most published studies on the development of prediction tools do not present regression coefficients and intercept of their newly developed models, rendering external validation by a third party impossible. We recommend that studies developing a prediction model report regression coefficients and intercepts in their manuscript enabling others to execute external validation.

Clinical applicability

Clinical applicability is one of the key aspects of a clinical useful prediction model, and should be kept in mind throughout the whole development and validation process. Clinical applicability already starts when selecting predictors. Unfortunately, there frequently exists a trade-off between the number of predictors and specificity of predictors on one side and the clinical applicability on the other side. Prediction models that use specific predictors, such as laboratory data, generally take more time before they can be used. However, their performance is usually better than models that use readily available parameters, such as demographics and vital signs. Following these principles, clinical applicability depends on the setting in which these models are used. Because of time constraints models for severely ill patients in the ED, should be based on readily available parameters. Once a prediction model is considered valid, it should be presented so that it is easy to use by clinicians. For example, prediction models can be presented as score charts, or introduced in mobile applications, or be implemented in electronic patients' files. A more user-friendly interface will enhance its use.

Generalisability

Generalisability is a very important matter whether or not to use a prediction model in a specific ED setting. Whether a prediction model is generalizable to other settings depends on the input, i.e. the population used to develop the model, and on the outcome variable. For the CLEARED model we used a database constructed in a tertiary referral medical center. The population presenting at the ED therefore differs from the population at many other (mostly secondary) referral centers. This should be taken into consideration when applying the model in other ED settings.

The outcome variable of CLEARED is admission to the hospital. A disadvantage of the outcome 'admission' as opposed to e.g. mortality is that it is dependent on the clinician. It is a decision and not a calculable outcome based on fixed fact, but made by doctors based on multiple factors. Doctors in one hospital may be more inclined to admit a patient to the ward, based on many aspects, than in another hospital: there is no golden standard. The decision to admit is influenced by the clinical presentation of the patient, but also by hospital capacity, distance from the hospital to the patients home, gut-feeling and multiple other factors. The influence of healthcare professionals' behaviour is less important when utilizing a more fixed outcome value, e.g. mortality. It is always important to validate a prediction tool in a new setting, before it is going to be used.

General recommendations

Based on our experience and research developing CLEARED and reviewing prediction models, we present three general recommendations with respect to prediction model development and evaluation.

First of all, while many research groups set out to develop their own prediction model from square one, this is ill advised as it ignores existing knowledge. Methodologically it is much more sound to execute external validation and update an existing model with additional parameters[353]. We, for example, added a measure of frailty and age to the NEWS, increasing its predictive power in an older population.

Secondly, local calibration and validation is of importance before implementation of a prediction tool. Ignoring these steps can overestimate the predictive value of a model, potentially harming individual patients.

Finally, after implementation prediction models should be evaluated with an impact analysis, a step that is hardly ever performed.

LIMITATIONS

The studies presented in this thesis have some limitations. The study designs were mostly retrospective and observational in nature. As a consequence, only associations of phenomena can be found and it is not possible to imply causality.

Another limitation with regards to studies with retrospectively gathered data is that analyses were limited to the available data. The data registered consisted of vital parameters, triage information and follow-up data. For example, the data in our studies consisted of parameters documented during the course of clinical diagnosis and treatment in patients. Data were not collected specifically for research, and therefore we may have excluded potentially important parameters of which the clinical relevance is not known. An associated limitation is the potential documentation errors. As most values were entered manually by nurses, errors due to inaccurate registration cannot be excluded. Most likely these errors occur at random. As it is not likely that these documentation errors are associated with outcomes, we do not expect this limitation to have a significant impact on our findings.

The restrictions on the available data in our studies poses an additional limitation. The parameters were recorded at one point in time at ED entry as opposed to serial record-

ings throughout the ED stay. Deterioration, stabilisation or improvement of consecutive measurements may be of more prognostic value than the initial value itself. Regrettably these changes in the vital parameters in the ED are not included in the analysis as these data were not available. On the other hand, the retrospective inclusion of an unselected patient cohort with routinely collected data has provided a large sample size and ensures generalizability.

RECOMMENDATIONS FOR FUTURE RESEARCH

The studies in this thesis form a foundation for implementation and evaluation of prediction tools in the ED, to improve the assessment and treatment of ED patients. However, future research should be executed to expand the use of prediction tools. More explicitly, research should focus on the implementation of CLEARED, improved recognition of deterioration in specific patient populations, the applicability of machine-learning techniques, and POCT in prediction tools.

Implementation and prospective evaluation of CLEARED

The CLEARED tool has the potential to significantly shorten the LOS of older patients in the ED, which is beneficial for both individual patients and patient groups. Median ED LOS in Erasmus MC is two hours and fifty minutes (IQR 1.53 - 4.05 hours) and there is a trend of an increasing LOS over the years. Whether the use of CLEARED actually shortens LOS at the ED and prevents adverse outcomes should be investigated prospectively.

The database we used contains data from all patients presenting at the ED. Therefore, we are able to expand the CLEARED tool to be applicable to the whole ED population. The presently gained insight in predictors of admission and mortality, can be used in the prospective development of a prediction model. Prospective data collection will be beneficial to prediction model performance, as it will reduce the number of missing values and will minimize potential confounding[78].

Implementation of CLEARED should be performed in a multicentre study in which multiple hospitals implement the prediction tool in a step-wedge design to account for seasonal variability[354]. Eventually, even secondary outcomes such as in hospital LOS can be studied. In order to refine and improve the validity of the model, a model should be re-evaluated regularly, preferably in large studies.

Assessment of illness in the older ED population

Considering that the older ED population is more at risk for adverse outcomes, future research should also focus on recognizing patients with serious illnesses that are in need of urgent interventions as early as possible. Current screeners and prediction models used in a general ED population, insufficiently recognize disease in older patients. More population-based studies are needed to determine better cut-off values of vital signs in older patients for subsequent inclusion in a geriatric EWS. As clinical presentation in older patients may be different, non-specific other "vital" signs, such as mobility and self-reported pain assessment, should also be considered for inclusion in EWS. Furthermore, part of the older population is frail before entering the ED, and has an even higher risk of adverse outcomes. Identifying the frail patient in the ED would help to alert clinicians that a different interpretation of signs and symptoms and an early start of treatment is warranted. In this way undertriage of older patients, as is frequently happening nowadays, can be prevented. Currently, existing geriatric screeners are time-consuming and require trained staff such as a geriatrician. As such, they do not speed up the process of evaluating the patient at the ED. Future studies should investigate whether combining EWS with simple geriatric screeners, or triage systems with geriatric specific components improve recognition of acuity in older patients, potentially leading to less undertriage and better triage.

Incorporation of machine learning for prediction

Prediction models may benefit from the upcoming techniques in artificial intelligence. Software is being developed that makes computers capable of machine learning. Computers are able to analyse data, learn from that data and apply the learned matter to make decisions[355]. Artificial intelligence is already incorporated in several prediction tools. *Hong et al.* provided a prediction tool for admission based on 972 parameters using machine learning techniques[356]. While the AUC of 0.92, was considerably higher than the AUC of our CLEARED tool, this model lacked external validation.

Machine-learning also has several potential drawbacks. A disadvantage of machine learning is that external validation by a third party is very difficult, as it is not transparent. External validation of machine learning models is extremely important as they often include large numbers of predictors and complex relationships and thus are very prone to overfitting[357]. Secondly, machine learning does not take pathophysiological mechanisms into account, potentially finding irrelevant association between the outcome and otherwise insignificant factors. Thirdly, a prediction model that includes many parameters requires to be embedded in electronic health records, as entering all these variables manually is not feasible. It is also difficult to imagine that this amount of data on a single patient can be collected shortly after entering the ED. These considerations may

make such models less clinically applicable and generalizable. Implementation of these models in different hospitals can also be cumbersome since not all hospitals utilizes electronic patient files and if they do, they do not use the same electronic health system.

In theory every observation, measurement or test contains information and could therefore contribute to predictions. This knowledge can be used to make highly detailed prediction models which can be specifically adapted to a certain hospital, to serve a hospital specific population.

Nevertheless, as time progresses and technology advances the potential of artificial intelligence and its applicability most likely will grow. In future, machine-learning can probably aid decision-making at the ED, enabling validation and calibration on a large scale. However, much work has to be done to make prediction models primarily based on readily available data collected from patients at ED entrance.

Speeding up the process in the ED

The time needed to get the results from diagnostics contributes to the total time spent in the ED. The results of conventional laboratory tests and radiological imaging can take more than one hour and thus limit a fast disposition of patients. POCT, POCUS and Fast track systems have a great potential of accelerating the diagnostic and therapeutic process.

POCT

Several studies have been published studying POCT in the ED. *Singer et al.* showed that early POCT at triage reduced ED LOS by approximately one hour[201]. *Chaisirin et al.* showed a reduction of time to decision-making (98 minutes) and a reduction in ED LOS (155,5 minutes) when using POCT[358]. A survey amongst acute internists revealed that POCT laboratory testing is available in the EDs of most teaching hospitals and tertiary care centers in the Netherlands. However, structural use is not implemented. A study to elucidate the value of POCT as (partial) replacement of routine laboratory testing in the ED is indispensable.

POCT can also be used in prediction models and triage systems. POCT parameters could be expanded in prediction models and eventually be implemented as a 'retriage-tool', an adjunct to current triage systems. Such a prediction tool in which POCT is incorporated could possibly lead to earlier identification of high-risk patients for whom the diagnostic tract should be further expedited and low-risk patients that can safely be discharged. Another goal of a POCT-based prediction tool is that it can help to create alternative patient flows, for example earlier admission to an AMU awaiting full assessment by the physician. Alternatively, bed assignment can be advanced in the ED process. With the

bed being prepared earlier, the patient can be transferred to the ward sooner. To study such interventions a multicentre study in which we can implement standard use of POCT in a step-wedged design is preferred.

POCUS

Next to POCT, point-of-care ultrasound (POCUS) as an extension of physical examination could also reduce ED LOS in non-trauma patients as it reduces the waiting time for a radiologist to perform an ultrasound[359, 360]. Together with the POCT the effect of POCUS should be evaluated prospectively to properly give both instruments of clinical evaluation, their proper place at the ED. Its value alone or in combination with other interventions also needs studying.

Fast track systems

Fast track systems in the ED are mostly implemented for low acuity patients with minor injuries or benign medical conditions[361, 362]. During the COVID-19 pandemic, where Dutch hospitals experienced large numbers of COVID-19 patients[363], several hospitals organised the admission process of COVID-19 patients as a fast track process. All essential diagnostics were done within 30 minutes after a patient entered the ED. Subsequently, patients were transported to the ward or sent home, based on admission criteria[364]. From the experience obtained from this fast track COVID-19 procedure it should be possible to create other fast track procedures in combination with POCT or POCUS. This should be aimed at groups at high risk for mortality such as older patients or patients suspected for sepsis.

CONCLUSION

This thesis is among the first PhD theses in acute medicine in the Netherlands. In this thesis, my aim was to predict with patients were at highest risk for admission and mortality independent of a specific disease and irrespective of underlying comorbidities. I developed, evaluated and validated prediction models for use in non-trauma ED patients that allow swift recognition of patients with the highest risk for admission and mortality, so that optimal treatment can be given as soon as possible and their stay in the ED can be shortened. The models and predictors reported in this thesis performed well in general and can be used to guide patient care. In elderly patients we should be aware of a different clinical presentation and older patients specific EWS should be developed. Future studies should focus on validation and extension of existing models, on the application of artificial intelligence and inclusion of POCT in prediction models. This will further improve clinical decision-making in the ED.


chapter
13

English Summary

Chapter 1 is a general introduction on the different topics in this thesis. After a description of the status of the emergency department (ED) within the Dutch healthcare system together with the history and use of triage systems, the development of valid prediction models and its most important features are described. Subsequently, we describe the usability of early warning scores and vital signs in the assessment of patients with potentially life-threatening diseases. Two of these illnesses described in this thesis are sepsis and hypertensive crises (i.e. an extremely elevated blood pressure). To conclude this introductory chapter a concise description of the databases we used is given.

This thesis is divided in four parts. The **first part** containing chapter 2 and 3 describes prediction models for admission of patients from the ED to the hospital.

In **chapter 2** a systematic review was conducted wherein several prediction models for admission of patients visiting the ED were evaluated. These models are supposed to be used to assess the chance for admission. When this chance is taken into account to decide whether a patients is to be admitted to the hospital this could markedly streamline the admission process, reduce length of stay at the ED and subsequently minimize crowding. Eleven studies in European EDs were included and sixteen different models were tested. The discriminative value varied between 0.63 and 0.88. Calibration was adequately used in just five studies. The developmental quality of the models was generally good. Eventually, we considered the models constructed by *Lucke et al.* and *Cameron et al.* to be the best as they had a well enough performance, were validated and used immediately available parameters. These models have not been implemented at the ED as of yet. Further research is needed to assess the applicability and implementation of admission models in the ED.

Chapter 3 describes the development of a model to predict admission in older ED patients. A prolonged stay in the ED is associated with a longer stay at the ward. It is also associated with development of a delirium, increased morbidity and even death. A good prediction of which patients are going to be admitted, could help to move patients much earlier to an inpatient ward far before all test results are known. This can be beneficial to individual patients since waiting time is reduced, but also to the overall process at the ED as it can lead to a reduction of throughput times and crowding. This model was called the CLEARED tool. We developed the CLEARED tool in the Erasmus MC. We used easily obtainable parameters like vital parameters, triage category, mode of transportation and the need to perform laboratory and radiology testing. We validated this tool internally and externally in two other hospitals. It had good discriminative and calibration properties and is thus able to predict admission. However, further additional

investigations are needed to implement the tool and to evaluate the effect on length of stay (LOS) and crowding in the ED, length of hospitalization and mortality.

The **second part** of this thesis consists of studies on predictors and prediction models for mortality. This part comprises of chapter 4 to chapter 7. We tested prediction models at the emergency department, both in the general ED population as in more specific populations, such as older patients and patients suffering from infection.

In **chapter 4** we systematically reviewed available literature to find existing prediction models aiming to predict mortality of patients presenting at the ED. We searched for models which were developed or validated in the ED of an European hospital. Seventeen studies were included, in which 22 prediction models were described. The area under the curve, a measure to differentiate between high and low risk of mortality, varied between 0.63 to 0.93. Many of these models were not calibrated, which reduces the reliability of them. Almost all models had an average risk of bias, which could lead to even more unreliable results. The two best models, combining the best model performance with the lowest risk of bias, were the Paris model and the full model by *Brabrand et al.* As of today these models have not yet been implemented in clinical practice. Prediction models could be useful to identify the more critical patients, in order to be able to treat these patients first. We concluded that the methodology to develop these models should be improved in order to develop a more reliable model. The yet existing models should be externally validated in a large dataset enabling a more reliable comparison of multiple prediction models.

In **chapter 5** we studied several prediction scores in their ability to predict mortality in patients presenting at the emergency department with a suspected infection. The Sepsis-3 definition led to the introduction of a new score, qSOFA. The tools tested were the "quick Sepsis Related Organ Failure Assessment" (qSOFA), the "Systemic Inflammatory Response Syndrome" (SIRS) and the "National Early Warning Score" (NEWS). NEWS proved to give the best predictions on both 10 day - and 30 day mortality, followed by qSOFA and SIRS. The calibration of NEWS was also superior compared to the other two tools. We found that qSOFA had the highest sensitivity of the three tools, while SIRS had the highest specificity. NEWS was the moderate option for both sensitivity and specificity. These results indicate that NEWS is the most suitable for predicting mortality at the ED, although the score was not intended to be used for patients specifically suffering from sepsis. More research is needed to confirm these results.

In **chapter 6** we tested EWS performance in a population of older patients visiting the ED. It was suspected that these scores were not suitable for patients older than seventy

years as the normal values of vital parameters change in aging people. The National Early Warning Score (NEWS), Modified Early Warning Score (MEWS), Rapid Acuity Physiology Score (RAPS) and the Rapid Emergency Medicine Score (REMS) were tested. This yielded mediocre results. Refitting the NEWS affirmed that the predictors have a different predictive value in older patients than in the general population. All four scores were evaluated in the cohort of the "acutely presenting older patients" (APOP) study, which was used in earlier studies to develop the APOP screener. This screener aims to predict adverse outcomes, especially functional loss, in order to be able to intervene in high risk patients. In this study we show that combining the APOP screener and the NEWS benefited the discriminative value. Adding age as an additional predictor increased this value even more. These findings suggest that in order to make the use of EWS of more value in older patients, it should be either adjusted or an geriatric EWS should be developed.

Chapter 7 contains a study validating an already existing model which predicts mortality based on six laboratory tests and age. These laboratory tests are virtually always ordered for patients treated at the ED by internists. These tests were the serum sodium, urea, glucose, haemoglobin, the white cell count and platelet count. The discriminative value of this model turned out to be decent, yet the calibration was of mediocre quality. The drawback of using laboratory results in a prediction model however, is that waiting for the test results frequently takes up to one hour which could result in a considerable delay in treatment. This reduces the applicability of said prediction model. Point-of-care-testing can circumvent this disadvantage. For a limited number of tests point-of-care-testing provides a result within mere minutes. We reduced the existing model to a model in which only tests that can be done using point-of-care testing were included. This yielded very similar results to the previous model. In this the potential of point-of-care testing in prediction modelling is clearly demonstrated.

The **third part** of this thesis consists of four studies on drug prescription and adherence in the emergency department.

Chapter 8 aims to observe the adherence to guidelines for prescription of antibiotics in the ED. A database, containing all patient of whom a blood culture taken at the ED turned out to be positive, was used. Adherence to local and national guidelines for prescription of antibiotics by the physicians was studied. We furthermore assessed whether non-adherence to the guidelines led to overtreatment, undertreatment or an equivalent treatment. Only 43% of patient were treated according to protocol. Both under- and overtreatment were not associated with an increase in mortality. A lower degree of disease severity, was associated with undertreatment. We conclude that overtreatment

is not always justified, as no decrease in mortality is shown in this study. Overtreatment can furthermore lead to increased bacterial resistance to antibiotics.

In **chapter 9** the value of body temperature as a parameter to predict disadvantageous outcomes and antibiotic prescription was assessed. Fever and hypothermia can both be indicators of an ongoing infection, therefore early warning scores attribute a higher risk of deterioration to patients presenting with hyper- and hypothermia. The association of body temperature on both mortality and on the decision to prescribe antibiotics in the ED was investigated. The study was conducted in patients suspected of infection (defined as having blood culture samples taken). We found that patients being normothermic were less frequently treated with antibiotics than hypo- or hyperthermic patients. However, mortality was higher in normothermic patients. This association was also present in a sub selection of patients with positive blood cultures. Based on our findings we recommend that when the suspicion of an infection arises, antibiotic therapy should not be omitted or delayed only because fever is absent.

Chapter 10 describes a study into the association between antibiotics administered at the ED and the mortality of patients suffering from life-threatening infections. We aimed to test whether prescription of appropriate antibiotics is associated with a reduction of mortality. Previously conducted studies provide mixed results. Finding an explanation for these mixed results was the secondary aim of this study.

We could not find a lower mortality rate in patients treated with appropriately treated antibiotics when analysing all patients suffering from systemic infections, even after correction for confounders. When analysing a homogenous group of patient only treated with just one antibiotic agent, adequate therapy appeared to reduce mortality as compared to inadequate therapy. This shows that correction is not always possible when subject groups differ from each other. A homogenous research population is therefore preferred.

Chapter 11 concerns ED patients presenting with an extremely high blood pressure (i.e. hypertensive crisis). Hypertensive crises consist of hypertensive urgencies and hypertensive emergencies. Hypertensive emergencies are characterised by target organ damage caused by high blood pressure in contrast to hypertensive urgencies. In this chapter we describe both a retrospective as a prospective study regarding this subject.

We retrospectively assessed whether patients presenting with severe hypertension at the ED fulfilled the criteria for hypertensive emergency or hypertensive urgency. Also possible causes for these highly elevated blood pressures were evaluated. In 22.1% of

patients (impending) target organ damage was present. 30.6% of cases fulfilled the criteria for hypertensive urgency. In only 6% of cases non-adherence to treatment was reported as cause for high blood pressure. However, these results were expected to be higher, suggesting that physicians do not always assess adherence to existing blood pressure lowering therapy.

In a prospective study the adherence to previously prescribed antihypertensive medication in 53 patients with severe hypertension was determined. Patients who initially admitted to be non-adherent to medication treatment were excluded. The blood concentration of several types of antihypertensive medication that these patients were supposed to use were measured. 22.6% of patients who didn't admit to be non-adherent to medication, proved to be non-adherent to at least one type of their prescribed medications. Non-adherence to therapy proves to be a common problem in patients with hypertensive urgencies and emergencies. At the ED non-adherence should be considered more often as a cause for these conditions.

In the **final part** all previous study results are summarised and discussed.

Chapter 12 contains a discussion on our findings and compare them to other studies conducted within the field of internal acute medicine. This chapter concludes with prospects to future research and the general conclusion of this thesis. **Chapter 13** contains the English summary of this thesis. The Dutch summary is written in **chapter 14.**

chapter 14

Nederlandse Samenvatting

In **hoofdstuk 1** wordt een algemene introductie gegeven over de onderwerpen die in dit proefschrift beschreven staan. Het begint met de opzet van de spoedeisende hulp in Nederland, gevolgd door een beschrijving van de geschiedenis en toepasbaarheid van triagesystemen. Hierna volgt een beschrijving over de belangrijke kenmerken van een valide predictiemodel. Hierin worden de begrippen discriminatie, het vermogen om hoog risico groepen van laag risico groepen te onderscheiden en kalibratie, de mate van overeenstemming tussen de voorspelde kans en de geobserveerde kans, uitgelegd. Vervolgens wordt beschreven hoe een dergelijk predictiemodel op de spoedeisende hulp zou kunnen worden toegepast. Aangezien er in dit proefschrift in een aantal hoofdstukken gefocust wordt op sepsis en hypertensieve crisis, zullen deze ziektebeelden vooraf toegelicht worden.

Dit hoofdstuk eindigt met het doel van het proefschrift, de indeling en een beschrijving van de databases die zijn gebruikt om dit proefschrift tot stand te brengen.

Dit proefschrift is opgedeeld in vier delen. Het **eerste deel** bestaande uit hoofdstukken 2 en 3, beschrijft predictiemodellen voor opname vanaf de spoedeisende hulp.

In **hoofdstuk 2** wordt met een *systematic review* diverse voorspelmodellen beschreven voor ziekenhuisopname van patiënten die de spoedeisende hulp bezoeken. Deze modellen kunnen gebruikt worden om op individueel niveau het risico op opname te bepalen en daarmee eventueel de opname vanaf de spoedeisende hulp te bespoedigen, maar heeft ook de potentie om de zorg voor patiënten beter te stroomlijnen en daarmee mogelijk *crowding* te reduceren. In totaal werden er elf studies geïncludeerd, waarin zestien verschillende modellen werden getest. Het discriminerende vermogen varieerde tussen 0.63 en 0.88. Kalibratie werd in slechts vijf studies uitgevoerd, en bleek adequaat. De kwaliteit van de nieuw ontwikkelde modellen was over het algemeen goed. Uiteindelijk vonden we de modellen van *Lucke et al.* en *Cameron et al.* het beste, gezien deze modellen een goede performance hadden, goed waren gevalideerd en gebruik maakten van direct beschikbare parameters. Echter, deze modellen zijn beiden nog niet geïmplementeerd op de spoedeisende hulp. In conclusie dragen wij aan om deze modellen te implementeren in ziekenhuizen, waarbij we adviseren om eerst een externe validatie uit te voeren om te kijken of het model ook goed werkt in dat specifieke ziekenhuis.

Hoofdstuk 3 beschrijft de ontwikkeling en validatie van een model om ziekenhuisopnames van ouderen vanaf de spoedeisende hulp te voorspellen. Langer verblijf op de spoedeisende hulp wordt geassocieerd met een verlengde opnameduur in het ziekenhuis, maar ook met andere negatieve uitkomsten, zoals het ontstaan van een

delier en zelfs mortaliteit. Het voorspellen van een opname, gevolgd door een interventie, zoals het direct opnemen in het ziekenhuis of opname op een acute opname afdeling heeft mogelijk een gunstig effect voor zowel de individuele patiënt als voor de patiëntenstroom op de spoedeisende hulp. Potentieel zou hiermee ook *crowding* gereduceerd kunnen worden. Het CLEARED instrument is ontwikkeld in het Erasmus MC en maakt gebruik van parameters die makkelijk te verkrijgen zijn bij aankomst op de spoedeisende hulp, zoals vitale parameters, triage categorie, wijze van aankomst op de spoedeisende hulp en de noodzaak tot het verrichten van laboratorium onderzoek en radiologisch onderzoek. Dit voorspelmodel is vervolgens zowel intern als extern gevalideerd met daarbij goede discriminatie en goede kalibratie. Echter, voor invoering is een prospectieve evaluatie van het effect op de opnameduur en op de uitkomst gewenst.

Het **tweede deel** bestaat uit artikelen die gaan over voorspelmodellen van mortaliteit. Dit omvat hoofdstuk 4 tot en met hoofdstuk 7, waarbij voorspelmodellen in zowel de algemene spoedeisende hulp populatie worden ontwikkeld en getest, als in specifieke populaties, zoals patiënten met een infectie en oudere patiënten.

In **hoofdstuk 4** staat een overzicht van modellen die mortaliteit voorspellen van patiënten die zich op de spoedeisende hulp presenteren. De modellen werden met een *systematic review* van de literatuur gevonden en beschreven. Er werd specifiek gezocht naar modellen die op de spoedeisende hulp van een Europees ziekenhuis werden ontwikkeld of werden gevalideerd. In totaal werden 17 artikelen geïncludeerd, waarin 22 modellen werden beschreven. Het merendeel van deze studies keek naar mortaliteit in het ziekenhuis. Het vermogen om de hoog-risico van de laag-risico populatie te onderscheiden, uitgedrukt als de *area under the curve*, varieerde tussen 0.63 en 0.93. Echter, kalibratie van deze modellen ontbrak frequent. Tevens bleek bij analyse van de kwaliteit van de ontwikkeling van de modellen, dat vrijwel elk model een gemiddeld risico op bias had. De twee beste modellen, gedefinieerd als de modellen met de beste performance en het laagste risico op bias, waren het Paris model en het full model van *Brabrand et al.*. Echter deze modellen zijn nog niet geïmplementeerd op de spoedeisende hulp. Deze modellen zouden gebruikt kunnen worden om zieke patiënten spoediger te kunnen identificeren en vervolgens te behandelen. De conclusie van dit hoofdstuk is dat de methodologische kwaliteit zou moeten worden verbeterd om een beter model te ontwikkelen. Tevens pleiten we voor het onderzoeken van huidige modellen in een grote dataset, zodat er een betere vergelijking gemaakt zou kunnen worden.

Hoofdstuk 5 beschrijft een studie waarin patiënten die op de spoedeisende hulp kwamen vanwege een verdenking op een infectie werden onderzocht. Er werden drie scores getest in hun vermogen om mortaliteit te voorspellen, namelijk de *quick Sepsis Related*

Organ Failure Assessment (qSOFA), het *Systemic Inflammatory Response Syndrome* (SIRS) en de *National Early Warning Score* (NEWS). NEWS had het beste discriminerende vermogen om zowel tien- als dertig dagen mortaliteit te voorspellen, gevolgd door qSOFA en SIRS. Tevens was NEWS als enige goed gekalibreerd. In dit onderzoek bleek dat de qSOFA de hoogste sensitiviteit had, ten opzichte van SIRS, die de hoogste specificiteit had. NEWS zit hier precies tussenin. Onze resultaten lijken erop te wijzen dat NEWS het meest geschikt is voor gebruik op de spoedeisende hulp, hoewel deze score initieel niet ontwikkeld is voor sepsis patiënten specifiek. Vervolgonderzoeken naar de waarde van de NEWS op de spoedeisende hulp zouden moeten volgen.

In **hoofdstuk 6** worden diverse *early warning scores* getest in onder ouderen op de spoedeisende hulp patiënten. Het vermoeden bestond dat deze scores niet optimaal zouden werken, omdat vitale functies en de mate van ontregeling daarvan met de leeftijd veranderen. De bestudeerde scores waren de *National Early Warning Score* (NEWS), *Modified Early Warning Score* (MEWS), *Rapid Acuity Physiology Score* (RAPS) en de *Rapid Emergency Medicine Score* (REMS). De scores presteerden allen matig, waarbij we aantoonden dat de parameters in ouderen andere voorspellende waarden hadden. Voor dit onderzoek werd het "Acuut Presenterende Oudere Patiënt" (APOP) cohort gebruikt, waarin eerder een APOP screeningsinstrument in ontwikkeld werd. Deze screeningsinstrument heeft als doel om negatieve uitkomsten in ouderen te voorspellen en daar beter op in te kunnen spelen. We toonden aan dat het discriminerend vermogen verbeterde van de NEWS als we hieraan de uitkomst van de APOP screeningsinstrument toevoegden. De toevoeging van leeftijd als losse voorspeller verbeterde het discriminerend vermogen nog meer. De bevindingen suggereren dat de *early warning scores* aangepast zouden moeten worden, zodat deze beter toepasbaar zijn bij oudere patiënten.

Hoofdstuk 7 beschrijft een studie waarbij we een bestaand model valideren dat mortaliteit voorspelt op basis van zes bloeduitslagen en de leeftijd. Deze laboratorium testen worden in vrijwel iedere patiënt afgenomen, die voor de interne geneeskunde op de spoedeisende hulp komt, namelijk serum natrium, serum ureum, serum glucose, het hemoglobinegehalte, het leukocyten-getal en het trombocyten-getal. De studie toonde aan dat de discriminatie van dit model goed was, maar dat de kalibratie beter zou kunnen. Het nadeel van laboratorium testen in een voorspelmodel is dat het verkrijgen van de bloeduitslagen tot wel een uur kan duren. Dit beperkt de bruikbaarheid van een dergelijk model voor op de spoedeisende hulp. Een oplossing hiervoor is *point-of-care testing*. Met deze methode worden bloeduitslagen binnen enkele minuten verkregen. Echter, dit is slechts mogelijk voor een beperkt aantal labtesten. Reductie van het originele model tot een *point-of-care testing* model resulteerde in een vergelijkbaar sterk model als het originele model. Hoewel dit model nog verder gevalideerd en eventueel

aangepast dient te worden, laat deze studie wel de potentie van *point-of-care testing* in predictiemodellen zien.

Het **derde deel** van dit proefschrift beschrijft drie studies over het niet-naleven van regels en richtlijnen op de spoedeisende hulp.

Hoofdstuk 8 beschrijft de resultaten van een onderzoek naar het niet naleven van antibiotica richtlijnen op de spoedeisende hulp. Voor dit onderzoek werd er gebruik gemaakt van een database met positieve bloedkweken, die bij patiënten op de spoedeisende hulp waren afgenomen. Er werd retrospectief gekeken of de antibiotica richtlijnen werden nageleefd en of bij het niet naleven van deze richtlijn er sprake was van onderbehandeling, overbehandeling of equivalente behandeling. In deze studie kwam naar voren dat slechts 43% van de patiënten volgens de richtlijnen behandeld werd. Zowel onder- als overbehandeling leidde niet tot meer mortaliteit, maar een lagere ziekte-ernst was geassocieerd met onderbehandeling, terwijl een hogere ziekte-ernst geassocieerd was met overbehandeling. Met deze studie werd aangetoond dat overbehandeling niet gerechtvaardigd is, aangezien er geen grote overlevingswinst in deze populatie werd aangetoond. Tevens leidt overbehandeling tot antibioticaresistentie. Onderbehandeling was geassocieerd met minder antibiotische dekking vergeleken met behandeling volgens de richtlijnen.

In **hoofdstuk 9** werd gefocust op de lichaamstemperatuur. Koorts, maar ook ondertemperatuur is vaak een teken van een infectie. In *early warning scores* wordt standaard een hoger risico op achteruitgang gegeven aan patiënten met zowel een ondertemperatuur als koorts. Het doel van dit onderzoek was om te kijken wat de impact van lichaamstemperatuur is op mortaliteit en wat de impact is van lichaamstemperatuur op het voorschrijven en starten van antibiotica op de spoedeisende hulp. Het onderzoek werd uitgevoerd in patiënten die werden verdacht van het hebben van een infectie (gedefinieerd als het afnemen van bloedkweken op de spoedeisende hulp). Dit onderzoek liet zien dat mensen met een normale lichaamstemperatuur het minst vaak antibiotica voorgeschreven kregen. Echter, als we naar de mortaliteit kijken, is deze hoger voor mensen met een normale temperatuur ten opzichte van mensen met koorts. In een selectie van patiënten met positieve bloedkweken bleef deze associatie bestaan. Met deze studie werd aangetoond dat mortaliteit toeneemt bij een lager wordende temperatuur. Mogelijk wordt dit deels verklaard, doordat deze patiënten minder vaak antibiotica krijgen op de spoedeisende hulp. In toekomstige triage- en *early warning* systemen zou deze omgekeerde associatie van temperatuur en lichaamstemperatuur mee genomen moeten worden. Bovendien zou het wel of niet toedienen van antibiotica niet moeten afhangen van de lichaamstemperatuur.

In **hoofdstuk 10** wordt een onderzoek beschreven naar de associatie tussen het geven van antibiotica op de spoedeisende hulp en sterfte in patiënten met een potentieel levensbedreigende infectie. Wij wilden testen of het toedienen van adequate antibiotica geassocieerd is met een verlaging van de sterfte. Eerdere literatuur hiernaar toonde gemengde resultaten. Het tweede doel van dit onderzoek was om een verklaring voor deze gemengde resultaten te vinden. In de gehele populatie met een positieve bloedkweek kon geen lagere sterfte worden aangetoond, zelfs niet na het op diverse manieren corrigeren van verschillende factoren. Er bleek sprake te zijn van *residual confounding*. Als we vervolgens een homogene groep onderzochten, die allen slechts door een antibioticum werden behandeld bleek wel dat adequate therapie de mortaliteit verlaagd. Kortom, het is niet altijd mogelijk om te corrigeren voor alles als de te onderzoeken groepen erg van elkaar verschillen. Daarom zou bij het vergelijken van groepen, de meest homogene populatie gebruikt moeten worden.

In **hoofdstuk 11** wordt er gekeken naar patiënten die met een hypertensieve crisis zich presenteren op de spoedeisende hulp. Hypertensieve crises zijn onder te verdelen in hypertensieve urgenties en hypertensieve noodgevallen. Bij een hypertensief noodgeval is er sprake van eindorgaanschade op basis van de verhoogde bloeddruk in tegenstelling tot bij een hypertensieve urgentie. In dit hoofdstuk wordt zowel een retrospectieve als een prospectieve studie beschreven.

In het retrospectieve deel werd van alle patiënten met een hypertensieve crisis op de spoedeisende hulp bekeken of zij voldeden aan de criteria van een hypertensieve noodsituatie of urgentie, en wat mogelijke oorzaken zijn van de verhoogde bloeddruk. In 22.1% van de gevallen was er sprake van (dreigende) orgaanschade en in 30.6% van de gevallen bleek er sprake van hypertensieve urgentie. In slechts zes procent van de gevallen werd therapieontrouw voor bloeddrukverlagende medicijnen als oorzaak beschreven voor de verhoogde bloeddruk.

In het prospectieve onderzoek werd therapietrouw bestudeerd in 53 spoedeisende hulp patiënten met een hypertensieve crisis. De spiegel van diverse bloeddrukverlagende medicijnen werd in het bloed gemeten. Ondanks exclusie van patiënten die anamnestische therapieontrouw waren, bleek 22.6% van de patiënten therapieontrouw voor tenminste één van de bloeddrukverlagende medicijnen. Therapieontrouw is dus een veelvoorkomend probleem in patiënten met een hypertensieve- urgentie of noodgeval. Er moet op de spoedeisende hulp vaker rekening gehouden worden met therapieontrouw, omdat dit een specifieke benadering vereist.

Het **vierde en laatste deel** van het proefschrift bevat de discussie en samenvatting.

Uiteindelijk worden in **hoofdstuk 12** de bevindingen van de onderzoeken bediscussieerd en afgezet tegen ander onderzoek binnen het vakgebied van de acute interne geneeskunde. Dit hoofdstuk sluit af met perspectieven voor vervolgonderzoek, waarin onder meer wordt gesteld dat het gebruik van POCT in re-triage mogelijk een oplossing kan bieden voor *crowding* op de spoedeisende hulp. **Hoofdstuk 13** geeft een samenvatting van het proefschrift in het Engels. De Nederlandse samenvatting heeft u zojuist gelezen (**hoofdstuk 14**).

Appendices

References

REFERENCES

1. Schneider SM, Hamilton GC, Moyer P, Stapczynski JS. Definition of emergency medicine. Acad Emerg Med. 1998;5(4):348-51.
2. Gaakeer MI, van den Brand CL, Gips E, van Lieshout JM, Huijsman R, Veugelers R, et al. [National developments in Emergency Departments in the Netherlands: numbers and origins of patients in the period from 2012 to 2015] Landelijke ontwikkelingen in de Nederlandse SEH's. Ned Tijdschr Geneeskd. 2016;160:D970.
3. van der Linden MC, Lindeboom R, van der Linden N, van den Brand CL, Lam RC, Lucas C, et al. Self-referring patients at the emergency department: appropriateness of ED use and motives for self-referral. International Journal of Emergency Medicine. 2014;7(1):28.
4. Slockers MT, Zia Z, van Dijk T, van Laere I. Geef onverzekerde daklozen de zorg waar ze recht op hebben. Medisch Contact. 2019(33-34):40 - 3.
5. Kuster T, Nickel CH, Jenny MA, Blaschke LL, Bingisser R. Combinations of Symptoms in Emergency Presentations: Prevalence and Outcome. J Clin Med. 2019;8(3).
6. Coppes S, Veugelers R, Hessels RAPA, van den Brand CL, Gaakeer MI. Position within the hospital and role in the emergency department of emergency physicians in the Netherlands: a national survey. International Journal of Emergency Medicine. 2020;13(1).
7. Kremers MNT, Wachelder JJH, Nanayakkara PWB, Haak HR, Consortium OAMR. Organisation of internal medicine in acute care in the Netherlands: a detailed overview. Neth J Med. 2020;78(5):251-60.
8. Schouten IWEM, Dekker D, Diepeveen SHA. 'Spoedinternist' kan acute zorg versterken. Medisch Contact. 2013;68(14):734-6.
9. Weinerman ER, Rutzen SR, Pearson DA. Effects of Medical "Triage" in Hospital Emergency Service. Public Health Rep. 1965;80:389-99.
10. van der Wulp I, van Baar ME, Schrijvers AJ. Reliability and validity of the Manchester Triage System in a general emergency department patient population in the Netherlands: results of a simulation study. Emerg Med J. 2008;25(7):431-4.
11. van Ierland Y, van Veen M, Huibers L, Giesen P, Moll HA. Validity of telephone and physical triage in emergency care: The Netherlands Triage System. Fam Pract. 2011;28(3):334-41.
12. Mackway-Jones K, Marsden J, Windle J. Emergency Triage. 3rd ed. West Sussex: John Wiley & Sons, Ltd.; 2014. 192 p.
13. Radvinsky DS, Yoon RS, Schmitt PJ, Prestigiacomo CJ, Swan KG, Liporace FA. Evolution and development of the Advanced Trauma Life Support (ATLS) protocol: a historical perspective. Orthopedics. 2012;35(4):305-11.
14. Carrington M, Down J. Recognition and assessment of critical illness. Anaesthesia & Intensive Care Medicine. 2010;11(1):6-8.
15. Behind-the-Scenes: What to Expect During an ER Visit [Internet]: Baton Rouge General. 2019. [cited 6 December 2020]. Available from: https://www.brgeneral.org/news-blog/2019/august/behind-the-scenes-what-to-expect-during-an-er-vi/.
16. Crowding. Dallas: American College of Emergency Physicians, 2019 April 2019. Report No.
17. Hwang U, Concato J. Care in the emergency department: how crowded is overcrowded? Acad Emerg Med. 2004;11(10):1097-101.
18. Pines JM. Moving closer to an operational definition for ED crowding. Acad Emerg Med. 2007;14(4):382-3; author reply 3-4.

19. Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. Ann Emerg Med. 2008;52(2):126-36.
20. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA, Jr. A conceptual model of emergency department crowding. Ann Emerg Med. 2003;42(2):173-80.
21. Kawano T, Nishiyama K, Hayashi H. Execution of diagnostic testing has a stronger effect on emergency department crowding than other common factors: a cross-sectional study. PLoS One. 2014;9(10):e108447.
22. Forster AJ, Stiell I, Wells G, Lee AJ, van Walraven C. The effect of hospital occupancy on emergency department length of stay and patient disposition. Acad Emerg Med. 2003;10(2):127-33.
23. Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang LJ, Han W, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med. 2013;61(6):605-11 e6.
24. Ugglas Ba, Djärv T, Ljungman PLS, Holzmann MJ. Emergency department crowding associated with increased 30-day mortality: a cohort study in Stockholm Region, Sweden, 2012 to 2016. JACEP Open. 2020;1(6):1-8.
25. Wickman L, Svensson P, Djarv T. Effect of crowding on length of stay for common chief complaints in the emergency department: A STROBE cohort study. Medicine (Baltimore). 2017;96(44):e8457.
26. Schull MJ, Lazier K, Vermeulen M, Mawhinney S, Morrison LJ. Emergency department contributors to ambulance diversion: a quantitative analysis. Ann Emerg Med. 2003;41(4):467-76.
27. Atzema C, Bandiera G, Schull MJ. Emergency department crowding: the effect on resident education. Ann Emerg Med. 2005;45(3):276-81.
28. Bond K, Ospina MB, Blitz S, Afilalo M, Campbell SG, Bullard M, et al. Frequency, determinants and impact of overcrowding in emergency departments in Canada: a national survey. Healthc Q. 2007;10(4):32-40.
29. Nippak PM, Isaac WW, Ikeda-Douglas CJ, Marion AM, VandenBroek M. Is there a relation between emergency department and inpatient lengths of stay? Can J Rural Med. 2014;19(1):12-20.
30. van Galen LS, Lammers EM, Schoonmade LJ, Alam N, Kramer MH, Nanayakkara PW. Acute medical units: The way to go? A literature review. Eur J Intern Med. 2017;39:24-31.
31. Baztan JJ, Suarez-Garcia FM, Lopez-Arrieta J, Rodriguez-Manas L, Rodriguez-Artalejo F. Effectiveness of acute geriatric units on functional decline, living at home, and case fatality among older patients admitted to hospital for acute medical disorders: meta-analysis. BMJ. 2009;338:b50.
32. Lentz BA, Jenson A, Hinson JS, Levin S, Cabral S, George K, et al. Validity of ED: Addressing heterogeneous definitions of over-triage and under-triage. Am J Emerg Med. 2017;35(7):1023-5.
33. Storm-Versloot MN, Ubbink DT, Kappelhof J, Luitse JS. Comparison of an informally structured triage system, the emergency severity index, and the manchester triage system to distinguish patient priority in the emergency department. Acad Emerg Med. 2011;18(8):822-9.
34. Grossmann FF, Zumbrunn T, Ciprian S, Stephan FP, Woy N, Bingisser R, et al. Undertriage in older emergency department patients--tilting against windmills? PLoS One. 2014;9(8):e106203.
35. Grossmann FF, Zumbrunn T, Frauchiger A, Delport K, Bingisser R, Nickel CH. At risk of undertriage? Testing the performance and accuracy of the emergency severity index in older emergency department patients. Ann Emerg Med. 2012;60(3):317-25 e3.
36. Nickel CH, Nemec M, Bingisser R. Weakness as presenting symptom in the emergency department. Swiss Med Wkly. 2009;139(17-18):271-2.
37. Vogenberg FR. Predictive and prognostic models: implications for healthcare decision-making in a modern recession. Am Health Drug Benefits. 2009;2(6):218-22.

38. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-38.
39. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. JAMA. 2017;318(14):1377-84.
40. 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.
41. Hillman KM, Bristow PJ, Chey T, Daffurn K, Jacques T, Norman SL, et al. Duration of life-threatening antecedents prior to intensive care admission. Intensive Care Med. 2002;28(11):1629-34.
42. van Galen LS, Struik PW, Driesen BE, Merten H, Ludikhuize J, van der Spoel JI, et al. Delayed Recognition of Deterioration of Patients in General Wards Is Mostly Caused by Human Related Monitoring Failures: A Root Cause Analysis of Unplanned ICU Admissions. PLoS One. 2016;11(8):e0161393.
43. Morgan RJM, Williams F, Wright MM. An early warning scoring system for detecting developing critical illness. Clin Intensive Care. 1997;8:100.
44. National Early Warning Score (NEWS): Standardising the assessment of acute-illness severity in the NHS. Report of a working party. London: Royal College of Physicians, 2012.
45. Duncan H, Hutchison J, Parshuram CS. The Pediatric Early Warning System score: a severity of illness score to predict urgent medical need in hospitalised children. J Crit Care. 2006;21(3):271-8.
46. 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.
47. Singh A, Guleria K, Vaid NB, Jain S. Evaluation of maternal early obstetric warning system (MEOWS chart) as a predictor of obstetric morbidity: a prospective observational study. Eur J Obstet Gynecol Reprod Biol. 2016;207:11-7.
48. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM. 2001;94(10):521-6.
49. Weiss AJ, Wier LM, Stocks C, Blanchard J. Overview of Emergency Department Visits in the United States, 2011: Statistical Brief #174. 2006.
50. Ding R, Meng Y, Ma X. The Central Role of the Inflammatory Response in Understanding the Heterogeneity of Sepsis-3. Biomed Res Int. 2018;2018:5086516.
51. Finfer S. Life after sepsis - is it worth living? SMACC; Dublin2016.
52. Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. Chest. 1992;101(6):1644-55.
53. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Crit Care Med. 2003;31(4):1250-6.
54. Churpek MM, Zadravecz FJ, Winslow C, Howell MD, Edelson DP. Incidence and Prognostic Value of the Systemic Inflammatory Response Syndrome and Organ Dysfunctions in Ward Patients. Am J Respir Crit Care Med. 2015;192(8):958-64.
55. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801-10.
56. Yang WS, Kang HD, Jung SK, Lee YJ, Oh SH, Kim YJ, et al. A mortality analysis of septic shock, vasoplegic shock and cryptic shock classified by the third international consensus definitions (Sepsis-3). Clin Respir J. 2020;14(9):857-63.

57. Kim HI, Park S. Sepsis: Early Recognition and Optimized Treatment. Tuberc Respir Dis (Seoul). 2019;82(1):6-14.
58. Sweeney TE, Liesenfeld O, May L. Diagnosis of bacterial sepsis: why are tests for bacteremia not sufficient? Expert Rev Mol Diagn. 2019;19(11):959-62.
59. Gradel KO, Jensen US, Schonheyder HC, Ostergaard C, Knudsen JD, Wehberg S, et al. Impact of appropriate empirical antibiotic treatment on recurrence and mortality in patients with bacteraemia: a population-based cohort study. Bmc Infect Dis. 2017;17(1):122.
60. Whelton PK, Carey RM, Aronow WS, Casey DE, Jr., Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension. 2018;71(6):1269-324.
61. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies C. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet. 2002;360(9349):1903-13.
62. Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people. Lancet. 2014;383(9932):1899-911.
63. Patel KK, Young L, Howell EH, Hu B, Rutecki G, Thomas G, et al. Characteristics and Outcomes of Patients Presenting With Hypertensive Urgency in the Office Setting. Jama Internal Medicine. 2016;176(7):981-8.
64. Pan Y, Cai W, Cheng Q, Dong W, An T, Yan J. Association between anxiety and hypertension: a systematic review and meta-analysis of epidemiological studies. Neuropsychiatr Dis Treat. 2015;11:1121-30.
65. Zampaglione B, Pascale C, Marchisio M, CavalloPerin P. Hypertensive urgencies and emergencies - Prevalence and clinical presentation. Hypertension. 1996;27(1):144-7.
66. Crowding. Ann Emerg Med. 2019;74(3):e31.
67. Bernstein SL, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, et al. The Effect of Emergency Department Crowding on Clinically Oriented Outcomes. Academic Emergency Medicine. 2009;16(1):1-10.
68. Epstein SK, Huckins DS, Liu SW, Pallin DJ, Sullivan AF, Lipton RI, et al. Emergency department crowding and risk of preventable medical errors. Intern Emerg Med. 2012;7(2):173-80.
69. Jo S, Jeong T, Jin YH, Lee JB, Yoon J, Park B. ED crowding is associated with inpatient mortality among critically ill patients admitted via the ED: post hoc analysis from a retrospective study. American Journal of Emergency Medicine. 2015;33(12):1725-31.
70. Mataloni F, Pinnarelli L, Perucci CA, Davoli M, Fusco D. Characteristics of ED crowding in the Lazio Region (Italy) and short-term health outcomes. Intern Emerg Med. 2019;14(1):109-17.
71. Salehi L, Phalpher P, Valani R, Meaney C, Amin Q, Ferrari K, et al. Emergency department boarding: a descriptive analysis and measurement of impact on outcomes. Can J Emerg Med. 2018;20(6):929-37.
72. 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.
73. Lowthian JA, Curtis AJ, Cameron PA, Stoelwinder JU, Cooke MW, McNeil JJ. Systematic review of trends in emergency department attendances: an Australian perspective. Emerg Med J. 2011;28(5):373-7.

74. Paton A, Mitra B, Considine J. Longer time to transfer from the emergency department after bed request is associated with worse outcomes. Emerg Med Australas. 2019;31(2):211-5.
75. Ebker-White A, Bein KJ, Berendsen Russell S, Dinh MM. The Sydney triage to admission risk tool (START) to improve patient flow in an emergency department: a model of care implementation pilot study. BMC Emergency Medicine. 2019;19(1):79.
76. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Plos Med. 2009;6(7):e1000097.
77. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc. 2016;104(3):240-3.
78. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. Plos Med. 2014;11(10):e1001744.
79. Pencina MJ, D'Agostino RB. Evaluating Discrimination of Risk Prediction Models The C Statistic. Jama-J Am Med Assoc. 2015;314(10):1063-4.
80. Cameron A, Ireland AJ, McKay GA, Stark A, Lowe DJ. Predicting admission at triage: are nurses better than a simple objective score? Emergency Medicine Journal. 2017;34(1):2-7.
81. Cameron A, Rodgers K, Ireland A, Jamdar R, Mckay GA. A simple tool to predict admission at the time of triage. Emergency Medicine Journal. 2015;32(3):174-9.
82. Di Bari M, Salvi F, Roberts AT, Balzi D, Lorenzetti B, Morichi V, et al. Prognostic stratification of elderly patients in the emergency department: a comparison between the "Identification of Seniors at Risk" and the "Silver Code". J Gerontol A Biol Sci Med Sci. 2012;67(5):544-50.
83. Salvi F, Morichi V, Lorenzetti B, Rossi L, Spazzafumo L, Luzi R, et al. Risk stratification of older patients in the emergency department: comparison between the Identification of Seniors at Risk and Triage Risk Screening Tool. Rejuvenation Res. 2012;15(3):288-94.
84. Noel G, Bonte N, Persico N, Bar C, Luigi S, Roch A, et al. Real-time estimation of inpatient beds required in emergency departments. Eur J Emerg Med. 2019;26(6):440-5.
85. Zlotnik A, Alfaro MC, Perez MCP, Gallardo-Antolin A, Martinez JMM. Building a Decision Support System for Inpatient Admission Prediction With the Manchester Triage System and Administrative Check-in Variables. Cin-Comput Inform Nu. 2016;34(5):224-30.
86. Alam N, Vegting IL, Houben E, van Berkel B, Vaughan L, Kramer MH, et al. Exploring the performance of the National Early Warning Score (NEWS) in a European emergency department. Resuscitation. 2015;90:111-5.
87. Brouns SHA, Mignot-Evers L, Derkx F, Lambooij SL, Dieleman JP, Haak HR. Performance of the Manchester triage system in older emergency department patients: a retrospective cohort study. BMC Emerg Med. 2019;19(1):3.
88. Kraaijvanger N, Rijpsma D, Roovers L, van Leeuwen H, Kaasjager K, van den Brand L, et al. Development and validation of an admission prediction tool for emergency departments in the Netherlands. Emergency Medicine Journal. 2018;35(8):464-70.
89. Lucke JA, de Gelder J, Clarijs F, Heringhaus C, de Craen AJM, Fogteloo AJ, et al. Early prediction of hospital admission for emergency department patients: a comparison between patients younger or older than 70 years. Emergency Medicine Journal. 2018;35(1):18-27.
90. Centers for Disease Control and Prevention. National Hospital Ambulatory Medical Care Survey: 2011 Emergency Department Summary Tables 2014 [cited 2016 September 9]. Available from: http://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2011_ed_web_tables.pdf.
91. Downing A, Wilson R. Older people's use of Accident and Emergency services. Age Ageing. 2005;34(1):24-30.

92. 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.
93. van den Akker M, Buntinx F, Metsemakers JF, Roos S, Knottnerus JA. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. J Clin Epidemiol. 1998;51(5):367-75.
94. Hajjar ER, Cafiero AC, Hanlon JT. Polypharmacy in elderly patients. Am J Geriatr Pharmacother. 2007;5(4):345-51.
95. Casalino E, Wargon M, Peroziello A, Choquet C, Leroy C, Beaune S, et al. Predictive factors for longer length of stay in an emergency department: a prospective multicentre study evaluating the impact of age, patient's clinical acuity and complexity, and care pathways. Emerg Med J. 2014;31(5):361-8.
96. Kreindler SA, Cui Y, Metge CJ, Raynard M. Patient characteristics associated with longer emergency department stay: a rapid review. Emerg Med J. 2016;33(3):194-9.
97. Latham LP, Ackroyd-Stolarz S. Emergency department utilization by older adults: a descriptive study. Can Geriatr J. 2014;17(4):118-25.
98. Limpawattana P, Phungoen P, Mitsungnern T, Laosuangkoon W, Tansangworn N. Atypical presentations of older adults at the emergency department and associated factors. Arch Gerontol Geriatr. 2016;62:97-102.
99. Salvi F, Morichi V, Grilli A, Giorgi R, De Tommaso G, Dessi-Fulgheri P. The elderly in the emergency department: a critical review of problems and solutions. Intern Emerg Med. 2007;2(4):292-301.
100. Vanpee D, Swine C, Vandenbossche P, Gillet JB. Epidemiological profile of geriatric patients admitted to the emergency department of a university hospital localized in a rural area. Eur J Emerg Med. 2001;8(4):301-4.
101. Schrijver EJ, Toppinga Q, de Vries OJ, Kramer MH, Nanayakkara PW. An observational cohort study on geriatric patient profile in an emergency department in the Netherlands. Neth J Med. 2013;71(6):324-30.
102. Liew D, Liew D, Kennedy MP. Emergency department length of stay independently predicts excess inpatient length of stay. Med J Aust. 2003;179(10):524-6.
103. Hwang U, Morrison RS. The geriatric emergency department. J Am Geriatr Soc. 2007;55(11):1873-6.
104. Shankar KN, Bhatia BK, Schuur JD. Toward patient-centered care: a systematic review of older adults' views of quality emergency care. Ann Emerg Med. 2014;63(5):529-50 e1.
105. Barfod C, Lauritzen MM, Danker JK, Soletormos G, Forberg JL, Berlac PA, et al. Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency department - a prospective cohort study. Scand J Trauma Resusc Emerg Med. 2012;20:28.
106. Considine J, Thomas S, Potter R. Predictors of critical care admission in emergency department patients triaged as low to moderate urgency. J Adv Nurs. 2009;65(4):818-27.
107. American College of Surgeons Committee on Trauma. Advanced trauma life support course for physicians. American College of Surgeons; Chicago, IL1989.
108. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet. 1974;2(7872):81-4.
109. Stenhouse C, Coates S, Tivey M, Allsop P, Parker T. Prospective evaluation of a modified Early Warning Score to aid earlier detection of patients developing critical illenss on a general surgical ward. British Journal of Anaesthesia. 2000;84(5):663.
110. Steyerberg EW. Clinical Prediction Models. New York: Springer Science + Business; 2009. p. 130-1.

111. Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall; 1993. p. 1-10.
112. Steyerberg EW. Clinical Prediction Models. New York: Springer Science + Business; 2009. p. 356-7.
113. Miller ME, Langefeld CD, Tierney WM, Hui SL, McDonald CJ. Validation of probabilistic predictions. Med Decis Making. 1993;13(1):49-58.
114. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81.
115. Dijkland SA, Roozenbeek B, Brouwer PA, Lingsma HF, Dippel DW, Vergouw LJ, et al. Prediction of 60-Day Case Fatality After Aneurysmal Subarachnoid Hemorrhage: External Validation of a Prediction Model. Crit Care Med. 2016;44(8):1523-9.
116. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. [
117. R Core Team. foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ... 2014 [Available from: http://CRAN.R-project.org/package=foreign.
118. Harrell FE, Jr. rms: Regression Modeling Strategies 2014, R package version 4.2-1 [Available from: http://CRAN.R-project.org/package=rms.
119. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. URL: http://www.rstudio.com/. [
120. Barron EA, Holmes J. Delirium within the emergency care setting, occurrence and detection: a systematic review. Emerg Med J. 2013;30(4):263-8.
121. Bo M, Bonetto M, Bottignole G, Porrino P, Coppo E, Tibaldi M, et al. Length of Stay in the Emergency Department and Occurrence of Delirium in Older Medical Patients. J Am Geriatr Soc. 2016;64(5):1114-9.
122. McCusker J, Bellavance F, Cardin S, Trepanier S, Verdon J, Ardman O. Detection of older people at increased risk of adverse health outcomes after an emergency visit: the ISAR screening tool. J Am Geriatr Soc. 1999;47(10):1229-37.
123. Meldon SW, Mion LC, Palmer RM, Drew BL, Connor JT, Lewicki LJ, et al. A brief risk-stratification tool to predict repeat emergency department visits and hospitalizations in older patients discharged from the emergency department. Acad Emerg Med. 2003;10(3):224-32.
124. Baumann MR, Strout TD. Triage of geriatric patients in the emergency department: validity and survival with the Emergency Severity Index. Ann Emerg Med. 2007;49(2):234-40.
125. LaMantia MA, Platts-Mills TF, Biese K, Khandelwal C, Forbach C, Cairns CB, et al. Predicting hospital admission and returns to the emergency department for elderly patients. Acad Emerg Med. 2010;17(3):252-9.
126. Yip WL, Fan KL, Lui CT, Leung LP, Ng F, Tsui KL. Utilization of the Accident & Emergency Departments by Chinese elderly in Hong Kong. World J Emerg Med. 2015;6(4):283-8.
127. Steiner D, Renetseder F, Kutz A, Haubitz S, Faessler L, Anderson JB, et al. Performance of the Manchester Triage System in Adult Medical Emergency Patients: A Prospective Cohort Study. J Emerg Med. 2016;50(4):678-89.
128. Groarke JD, Gallagher J, Stack J, Aftab A, Dwyer C, McGovern R, et al. Use of an admission early warning score to predict patient morbidity and mortality and treatment success. Emerg Med J. 2008;25(12):803-6.
129. Tape T. The Area Under an ROC Curve. Interpreting Diagnostic Test, University of Nebraska Medical Center. [Available from: http://gim.unmc.edu/dxtests/ROC3.htm.
130. Lemeshow S, Hosmer DW, Jr. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol. 1982;115(1):92-106.

131. Brabrand M, Lassen AT, Knudsen T, Hallas J. Seven-day mortality can be predicted in medical patients by blood pressure, age, respiratory rate, loss of independence, and peripheral oxygen saturation (the PARIS Score): A prospective cohort study with external validation. PLoS ONE. 2015;10(4).
132. Coslovsky M, Takala J, Exadaktylos AK, Martinolli L, Merz TM. A clinical prediction model to identify patients at high risk of death in the emergency department. Intensive Care Med. 2015;41(6):1029-36.
133. 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. 2010;103(1):23-32.
134. Kellett J, Deane B. The Simple Clinical Score predicts mortality for 30 days after admission to an acute medical unit. QJM. 2006;99(11):771-81.
135. Kellett J, Deane B, Gleeson M. Derivation and validation of a score based on Hypotension, Oxygen saturation, low Temperature, ECG changes and Loss of independence (HOTEL) that predicts early mortality between 15 min and 24 h after admission to an acute medical unit. Resuscitation. 2008;78(1):52-8.
136. Stræde M, Brabrand M. External validation of the simple clinical score and the HOTEL score, two scores for predicting short-term mortality after admission to an acute medical unit. PLoS ONE. 2014;9(8).
137. Cournane S, Conway R, Byrne D, O'Riordan D, Silke B. Predicting Outcomes in Emergency Medical Admissions Using a Laboratory Only Nomogram. Comput Math Method M. 2017.
138. Kristensen M, Iversen AKS, Gerds TA, Ostervig R, Linnet JD, Barfod C, 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.
139. O'Sullivan E, Callely E, O'Riordan D, Bennett K, Silke B. Predicting outcomes in emergency medical admissions - Role of laboratory data and co-morbidity. Acute Med. 2012;11(2):59-65.
140. Slagman A, Searle J, Vollert JO, Müller C, Muller R, Somasundaram R, et al. Predicting in-hospital mortality using routine parameters in unselected nonsurgical emergency department patients. Notf Rettungsmed. 2015;18(6):501-9.
141. Duckitt RW, Buxton-Thomas R, Walker J, Cheek E, Bewick V, Venn R, et al. Worthing physiological scoring system: Derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study. Br J Anaesth. 2007;98(6):769-74.
142. Bulut M, Cebicci H, Sigirli D, Sak A, Durmus O, Top AA, et al. The comparison of modified early warning score with rapid emergency medicine score: A prospective multicentre observational cohort study on medical and surgical patients presenting to emergency department. Emerg Med J. 2014;31(6):476-81.
143. Olsson T, Terent A, Lind L. Rapid Emergency Medicine score: a new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. J Intern Med. 2004;255(5):579-87.
144. Goodacre S, Turner J, Nicholl J. Prediction of mortality among emergency medical admissions. Emerg Med J. 2006;23(5):372-5.
145. Merz TM, Etter R, Mende L, Barthelmes D, Wiegand J, Martinolli L, et al. Risk assessment in the first fifteen minutes: A prospective cohort study of a simple physiological scoring system in the emergency department. Crit Care. 2011;15(1).
146. Bech CN, Brabrand M, Mikkelsen S, Lassen A. Risk factors associated with short term mortality changes over time, after arrival to the emergency department. Scand J Trauma Resusc Emerg Med. 2018;26(1):29.

147. Altman DG, Bland JM. Missing data. BMJ. 2007;334(7590):424.
148. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925-31.
149. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-9.
150. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338:b375.
151. Seymour CW, Liu VX, Iwashyna TJ. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) (vol 315, pg 762, 2016). Jama-J Am Med Assoc. 2016;315(20):2237-.
152. Carneiro AH, Povoa P, Gomes JA. Dear Sepsis-3, we are sorry to say that we don't like you. Rev Bras Ter Intensiva. 2017;29(1):4-8.
153. Simpson SQ. New Sepsis Criteria: A Change We Should Not Make. Chest. 2016;149(5):1117-8.
154. Askim A, Moser F, Gustad LT, Stene H, Gundersen M, Asvold BO, et al. Poor performance of quick-SOFA (qSOFA) score in predicting severe sepsis and mortality - a prospective study of patients admitted with infection to the emergency department. Scand J Trauma Resusc Emerg Med. 2017;25(1):56.
155. Canet E, Taylor DM, Khor R, Krishnan V, Bellomo R. qSOFA as predictor of mortality and prolonged ICU admission in Emergency Department patients with suspected infection. J Crit Care. 2018;48:118-23.
156. Goulden R, Hoyle MC, Monis J, Railton D, Riley V, Martin P, et al. qSOFA, SIRS and NEWS for predicting inhospital mortality and ICU admission in emergency admissions treated as sepsis. Emerg Med J. 2018;35(6):345-9.
157. Hwang SY, Jo IJ, Lee SU, Lee TR, Yoon H, Cha WC, et al. Low Accuracy of Positive qSOFA Criteria for Predicting 28-Day Mortality in Critically Ill Septic Patients During the Early Period After Emergency Department Presentation. Ann Emerg Med. 2018;71(1):1-9 e2.
158. Royal College of Physicians. National Early Warning Score (NEWS): Standardising the assessment of acute illness severity in the NHS. Report of a working party. London: RCP; 2012.
159. Hoikka M, Silfvast T, Ala-Kokko TI. Does the prehospital National Early Warning Score predict the short-term mortality of unselected emergency patients? Scand J Trauma Resusc Emerg Med. 2018;26(1):48.
160. Kivipuro M, Tirkkonen J, Kontula T, Solin J, Kalliomaki J, Pauniaho SL, et al. National early warning score (NEWS) in a Finnish multidisciplinary emergency department and direct vs. late admission to intensive care. Resuscitation. 2018;128:164-9.
161. Alexander RH, Proctor HJ, American College of Surgeons, Committee on Trauma. Advanced Trauma Life Support Program For Physicians: ATLS. 5th ed. Chicago, IL: American College of Surgeons; 1993.
162. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017;9:157-66.
163. Jo S, Yoon J, Lee JB, Jin Y, Jeong T, Park B. Predictive value of the National Early Warning Score-Lactate for mortality and the need for critical care among general emergency department patients. J Crit Care. 2016;36:60-8.
164. Shaw J, Fothergill RT, Clark S, Moore F. Can the prehospital National Early Warning Score identify patients most at risk from subsequent deterioration? Emerg Med J. 2017.
165. Sbiti-Rohr D, Kutz A, Christ-Crain M, Thomann R, Zimmerli W, Hoess C, et al. The National Early Warning Score (NEWS) for outcome prediction in emergency department patients with

community-acquired pneumonia: results from a 6-year prospective cohort study. BMJ Open. 2016;6(9):e011021.

166. Bilben B, Grandal L, Sovik S. National Early Warning Score (NEWS) as an emergency department predictor of disease severity and 90-day survival in the acutely dyspneic patient - a prospective observational study. Scand J Trauma Resusc Emerg Med. 2016;24:80.
167. Corfield AR, Lees F, Zealley I, Houston G, Dickie S, Ward K, et al. Utility of a single early warning score in patients with sepsis in the emergency department. Emerg Med J. 2014;31(6):482-7.
168. Gill MR, Reiley DG, Green SM. Interrater reliability of Glasgow Coma Scale scores in the emergency department. Ann Emerg Med. 2004;43(2):215-23.
169. Semler MW, Stover DG, Copland AP, Hong G, Johnson MJ, Kriss MS, et al. Flash mob research: a single-day, multicenter, resident-directed study of respiratory rate. Chest. 2013;143(6):1740-4.
170. 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.
171. Sjoding MW, Luo K, Miller MA, Iwashyna TJ. When do confounding by indication and inadequate risk adjustment bias critical care studies? A simulation study. Crit Care. 2015;19:195.
172. Asayama K, Aikawa N. Evaluation of systemic inflammatory response syndrome criteria as a predictor of mortality in emergency patients transported by ambulance. Keio Journal of Medicine. 1998;47(1):19-27.
173. Jones GR, Lowes JA. The systemic inflammatory response syndrome as a predictor of bacteraemia and outcome from sepsis. QJM. 1996;89(7):515-22.
174. Park HK, Kim WY, Kim MC, Jung W, Ko BS. Quick sequential organ failure assessment compared to systemic inflammatory response syndrome for predicting sepsis in emergency department. J Crit Care. 2017;42:12-7.
175. Sprung CL, Schein RMH, Balk RA. The new sepsis consensus definitions: the good, the bad and the ugly. Intens Care Med. 2016;42(12):2024-6.
176. Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, et al. Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit. JAMA. 2017;317(3):290-300.
177. Churpek MM, Snyder A, Han X, Sokol S, Pettit N, Howell MD, et al. Quick Sepsis-related Organ Failure Assessment, Systemic Inflammatory Response Syndrome, and Early Warning Scores for Detecting Clinical Deterioration in Infected Patients outside the Intensive Care Unit. Am J Respir Crit Care Med. 2017;195(7):906-11.
178. Kause J, Smith G, Prytherch D, Parr M, Flabouris A, Hillman K, et al. A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom - the ACADEMIA study. Resuscitation. 2004;62(3):275-82.
179. Churpek MM, Yuen TC, Winslow C, Hall J, Edelson DP. Differences in vital signs between elderly and nonelderly patients prior to ward cardiac arrest. Crit Care Med. 2015;43(4):816-22.
180. de Groot B, Stolwijk F, Warmerdam M, Lucke JA, Singh GK, Abbas 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.
181. Samaras N, Chevalley T, Samaras D, Gold G. Older Patients in the Emergency Department: A Review. Annals of Emergency Medicine. 2010;56(3):261-9.
182. Bullard MJ, Musgrave E, Warren D, Unger B, Skeldon T, Grierson R, et al. Revisions to the Canadian Emergency Department Triage and Acuity Scale (CTAS) Guidelines 2016. CJEM. 2017;19(S2):S18-S27.

183. de Gelder J, Lucke JA, de Groot B, Fogteloo AJ, Anten S, Mesri K, et al. Predicting adverse health outcomes in older emergency department patients: the APOP study. Neth J Med. 2016;74(8):342-52.
184. de Gelder J, Lucke JA, Blomaard LC, Booijen AM, Fogteloo AJ, Anten S, 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.
185. Su TL, Jaki T, Hickey GL, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models. Stat Methods Med Res. 2018;27(1):185-97.
186. Norman DC. Fever and aging. Infect Dis Clin Prac. 1998;7(8):387-90.
187. Yao JL, Fang J, Lou QQ, Anderson RM. A systematic review of the identification of seniors at risk (ISAR) tool for the prediction of adverse outcome in elderly patients seen in the emergency department. Int J Clin Exp Med. 2015;8(4):4778-86.
188. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. Academic Emergency Medicine. 2000;7(3):236-42.
189. Geier F, Popp S, Greve Y, Achterberg A, Glockner E, Ziegler R, et al. Severity illness scoring systems for early identification and prediction of in-hospital mortality in patients with suspected sepsis presenting to the emergency department. Wien Klin Wochenschr. 2013;125(17-18):508-15.
190. van der Wulp I, Schrijvers AJ, van Stel HF. Predicting admission and mortality with the Emergency Severity Index and the Manchester Triage System: a retrospective observational study. Emerg Med J. 2009;26(7):506-9.
191. Griffiths JR, Kidney EM. Current use of early warning scores in UK emergency departments. Emergency Medicine Journal. 2012;29(1):65-6.
192. Tian L, Fang Z, Xiao H, Li L, Li Y. Value of triage early warning score for trauma patients in an emergency department. Zhong Nan Da Xue Xue Bao Yi Xue Ban. 2015;40(5):549-57.
193. Mitsunaga T, Hasegawa I, Uzura M, Okuno K, Otani K, Ohtaki Y, et al. Comparison of the National Early Warning Score (NEWS) and the Modified Early Warning Score (MEWS) for predicting admission and in-hospital mortality in elderly patients in the pre-hospital setting and in the emergency department. PeerJ. 2019;7:e6947.
194. Brink A, Alsma J, Fortuin AW, Bramer WM, Zietse R, Lingsma HF, et al. Prediction models for mortality in adult patients visiting the Emergency Department: a systematic review. Acute Med. 2019;18(3):171-83.
195. 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.
196. Asadollahi K, Hastings IM, Beeching NJ, Gill GV. Laboratory risk factors for hospital mortality in acutely admitted patients. Qjm-Int J Med. 2007;100(8):501-7.
197. Jung WJ, Lee HJ, Park S, Lee SN, Kang HR, Jeon JS, et al. Severity of community acquired hypernatremia is an independent predictor of mortality. Intern Emerg Med. 2017;12(7):935-40.
198. Waikar SS, Mount DB, Curhan GC. Mortality after Hospitalization with Mild, Moderate, and Severe Hyponatremia. Am J Med. 2009;122(9):857-65.
199. Whelan B, Bennett K, Oriordan D, Silke B. Serum sodium as a risk factor for in-hospital mortality in acute unselected general medical patients. Qjm-Int J Med. 2009;102(3):175-82.
200. Kankaanpaa M, Holma-Eriksson M, Kapanen S, Heitto M, Bergstrom S, Muukkonen L, et al. Comparison of the use of comprehensive point-of-care test panel to conventional laboratory process in emergency department. Bmc Emergency Medicine. 2018;18.

201. Singer AJ, Taylor M, LeBlanc D, Meyers K, Perez K, Thode HC, Jr., et al. Early Point-of-Care Testing at Triage Reduces Care Time in Stable Adult Emergency Department Patients. J Emerg Med. 2018;55(2):172-8.
202. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Jr., Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2019;38(7):1276-96.
203. Cross A, Levine MM. Patterns of bacteraemia aetiology. Lancet Infect Dis. 2017;17(10):1005-6.
204. Dellinger RP. The Surviving Sepsis Campaign: Where have we been and where are we going? Cleve Clin J Med. 2015;82(4):237-44.
205. Goto M, Al-Hasan MN. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. Clin Microbiol Infect. 2013;19(6):501-9.
206. Kollef MH. Broad-spectrum antimicrobials and the treatment of serious bacterial infections: getting it right up front. Clin Infect Dis. 2008;47 Suppl 1:S3-13.
207. Paterson DL. The role of antimicrobial management programs in optimizing antibiotic prescribing within hospitals. Clin Infect Dis. 2006;42 Suppl 2:S90-5.
208. Dellit TH, Owens RC, McGowan JE, Jr., Gerding DN, Weinstein RA, Burke JP, et al. Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America guidelines for developing an institutional program to enhance antimicrobial stewardship. Clin Infect Dis. 2007;44(2):159-77.
209. Marquet K, Liesenborgs A, Bergs J, Vleugels A, Claes N. Incidence and outcome of inappropriate in-hospital empiric antibiotics for severe infection: a systematic review and meta-analysis. Crit Care. 2015;19:63.
210. Paul M, Shani V, Muchtar E, Kariv G, Robenshtok E, Leibovici L. Systematic review and meta-analysis of the efficacy of appropriate empiric antibiotic therapy for sepsis. Antimicrob Agents Chemother. 2010;54(11):4851-63.
211. Galayduyk N, Colodner R, Chazan B, Flatau E, Lavi I, Raz R. Adherence to guidelines on empiric use of antibiotics in the emergency room. Infection. 2008;36(5):408-14.
212. Hagen TL, Hertz MA, Uhrin GB, Dalager-Pedersen M, Schonheyder HC, Nielsen H. Adherence to local antimicrobial guidelines for initial treatment of community-acquired infections. Dan Med J. 2017;64(6).
213. van der Velden LB, Tromp M, Bleeker-Rovers CP, Hulscher M, Kullberg BJ, Mouton JW, et al. Non-adherence to antimicrobial treatment guidelines results in more broad-spectrum but not more appropriate therapy. Eur J Clin Microbiol Infect Dis. 2012;31(7):1561-8.
214. CDC. Bloodstream Infection Event (Central Line-Associated Bloodstream Infection and Non-central Line Associated Bloodstream Infection) 2020 [Available from: https://www.cdc.gov/nhsn/pdfs/pscmanual/4psc_clabscurrent.pdf.
215. Trick WE, Zagorski BM, Tokars JI, Vernon MO, Welbel SF, Wisniewski MF, et al. Computer algorithms to detect bloodstream infections. Emerg Infect Dis. 2004;10(9):1612-20.
216. Zachariasse JM, Seiger N, Rood PP, Alves CF, Freitas P, Smit FJ, et al. Validity of the Manchester Triage System in emergency care: A prospective observational study. PLoS One. 2017;12(2):e0170811.
217. Shapiro NI, Wolfe RE, Wright SB, Moore R, Bates DW. Who needs a blood culture? A prospectively derived and validated prediction rule. J Emerg Med. 2008;35(3):255-64.
218. Friedman ND, Kaye KS, Stout JE, McGarry SA, Trivette SL, Briggs JP, et al. Health care--associated bloodstream infections in adults: a reason to change the accepted definition of community-acquired infections. Ann Intern Med. 2002;137(10):791-7.
219. McGinley A, Pearse RM. A national early warning score for acutely ill patients. BMJ. 2012;345:e5310.

220. Schuttevaer R, Brink A, Alsma J, De Steenwinkel JEM, Verbon A, Schuit SCE, et al. Datasets RePub, Erasmus University Repository - Medical microbiology database. 2020.
221. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. 2011;173(6):676-82.
222. Verbrugh HA. Mapping antibiotic use and resistance in the Netherlands: SWAB and NethMap. Neth J Med. 2003;61(11):341-2.
223. Bax HV, N.; Hunfeld, N. Het Antibioticaboekje van het Erasmus MC: Haas, P; 2018 [Available from: https://erasmusmc.swabid.nl/node/79221.
224. Hernán MAR, J.M. Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming. 2018.
225. Ewig S, Seifert K, Kleinfeld T, Goke N, Schafer H. Management of patients with community-acquired pneumonia in a primary care hospital: a critical evaluation. Respir Med. 2000;94(6):556-63.
226. Villar-Alvarez F, Moreno-Zabaleta R, Mira-Solves JJ, Calvo-Corbella E, Diaz-Lobato S, Gonzalez-Torralba F, et al. Do not do in COPD: consensus statement on overuse. Int J Chron Obstruct Pulmon Dis. 2018;13:451-63.
227. Schuttevaer R, Alsma J, Brink A, van Dijk W, de Steenwinkel JEM, Lingsma HF, et al. Appropriate empirical antibiotic therapy and mortality: Conflicting data explained by residual confounding. PLoS One. 2019;14(11):e0225478.
228. Gary NE, Buzzeo L, Salaki J, Eisinger RP. Gentamicin-associated acute renal failure. Arch Intern Med. 1976;136(10):1101-4.
229. Romanovsky AA, Almeida MC, Aronoff DM, Ivanov AI, Konsman JP, Steiner AA, et al. Fever and hypothermia in systemic inflammation: recent discoveries and revisions. Front Biosci. 2005;10:2193-216.
230. Nawar EW, Niska RW, Xu J. National Hospital Ambulatory Medical Care Survey: 2005 emergency department summary. Adv Data. 2007(386):1-32.
231. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3:17.
232. Brink A, Alsma J, Verdonschot R, Rood PPM, Zietse R, Lingsma HF, et al. Predicting mortality in patients with suspected sepsis at the Emergency Department; A retrospective cohort study comparing qSOFA, SIRS and National Early Warning Score. PLoS One. 2019;14(1):e0211133.
233. Rumbus Z, Matics R, Hegyi P, Zsiboras C, Szabo I, Illes A, et al. Fever Is Associated with Reduced, Hypothermia with Increased Mortality in Septic Patients: A Meta-Analysis of Clinical Trials. PLoS One. 2017;12(1):e0170152.
234. Stoneking LR, Winkler JP, DeLuca LA, Stolz U, Stutz A, Luman JC, et al. Physician documentation of sepsis syndrome is associated with more aggressive treatment. West J Emerg Med. 2015;16(3):401-7.
235. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Med. 2017;43(3):304-77.
236. Vincent JL. The Clinical Challenge of Sepsis Identification and Monitoring. Plos Med. 2016;13(5):e1002022.
237. Romanelli D, Farrell MW. AVPU (Alert, Voice, Pain, Unresponsive). 2020.
238. Prevention CfDCa. Bloodstream Infection Event (Central Line-Associated Bloodstream Infection and Non-central Line Associated Bloodstream Infection) 2019 [Available from: https://www.cdc.gov/nhsn/pdfs/pscmanual/4psc_clabscurrent.pdf.

239. 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.
240. Recently updated guidelines from the National Institute for Health and Care Excellence (NICE) and the royal colleges. BMJ. 2017;357:j1363.
241. O'Grady NP, Barie PS, Bartlett JG, Bleck T, Carroll K, Kalil AC, et al. Guidelines for evaluation of new fever in critically ill adult patients: 2008 update from the American College of Critical Care Medicine and the Infectious Diseases Society of America. Crit Care Med. 2008;36(4):1330-49.
242. Schuttevaer R, Brink A, Alsma J, van Dijk W, Melles DC, de Steenwinkel JEM, et al. Non-adherence to antimicrobial guidelines in patients with bloodstream infection visiting the emergency department. Eur J Intern Med. 2020;78:69-75.
243. Alam N, Oskam E, Stassen PM, Exter PV, van de Ven PM, Haak HR, et al. Prehospital antibiotics in the ambulance for sepsis: a multicentre, open label, randomised trial. Lancet Respir Med. 2018;6(1):40-50.
244. Anderson DJ, Moehring RW, Sloane R, Schmader KE, Weber DJ, Fowler VG, Jr., et al. Bloodstream infections in community hospitals in the 21st century: a multicenter cohort study. PLoS One. 2014;9(3):e91713.
245. Cain SE, Kohn J, Bookstaver PB, Albrecht H, Al-Hasan MN. Stratification of the impact of inappropriate empirical antimicrobial therapy for Gram-negative bloodstream infections by predicted prognosis. Antimicrob Agents Chemother. 2015;59(1):245-50.
246. Corona A, Bertolini G, Lipman J, Wilson AP, Singer M. Antibiotic use and impact on outcome from bacteraemic critical illness: the BActeraemia Study in Intensive Care (BASIC). J Antimicrob Chemother. 2010;65(6):1276-85.
247. Fitzpatrick JM, Biswas JS, Edgeworth JD, Islam J, Jenkins N, Judge R, et al. Gram-negative bacteraemia; a multi-centre prospective evaluation of empiric antibiotic therapy and outcome in English acute hospitals. Clin Microbiol Infect. 2016;22(3):244-51.
248. Savage RD, Fowler RA, Rishu AH, Bagshaw SM, Cook D, Dodek P, et al. The Effect of Inadequate Initial Empiric Antimicrobial Treatment on Mortality in Critically Ill Patients with Bloodstream Infections: A Multi-Centre Retrospective Cohort Study. PLoS One. 2016;11(5):e0154944.
249. Thom KA, Shardell MD, Osih RB, Schweizer ML, Furuno JP, Perencevich EN, et al. Controlling for severity of illness in outcome studies involving infectious diseases: impact of measurement at different time points. Infect Control Hosp Epidemiol. 2008;29(11):1048-53.
250. McGregor JC, Rich SE, Harris AD, Perencevich EN, Osih R, Lodise TP, Jr., et al. A systematic review of the methods used to assess the association between appropriate antibiotic therapy and mortality in bacteremic patients. Clin Infect Dis. 2007;45(3):329-37.
251. Hernán MA, Robins JM. Causal Inference: Boca Raton: Chapman & Hall/CRC, forthcoming; 2019.
252. Bax HV, N.; Hunfeld, N. SWAB Erasmus MC 2018 [Available from: https://erasmusmc.swabid.nl/.
253. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. Am J Epidemiol. 2006;163(12):1149-56.
254. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci. 2007;8(3):206-13.
255. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006;98(3):253-9.
256. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.
257. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Stat Med. 2016;35(30):5642-55.

258. Rassen JA, Glynn RJ, Rothman KJ, Setoguchi S, Schneeweiss S. Applying propensity scores estimated in a full cohort to adjust for confounding in subgroup analyses. Pharmacoepidemiol Drug Saf. 2012;21(7):697-709.
259. Karras DJ, Kruus LK, Cienki JJ, Wald MM, Chiang WK, Shayne P, et al. Evaluation and treatment of patients with severely elevated blood pressure in academic emergency departments: a multicenter study. Ann Emerg Med. 2006;47(3):230-6.
260. Patel KK, Young L, Howell EH, Hu B, Rutecki G, Thomas G, et al. Characteristics and Outcomes of Patients Presenting With Hypertensive Urgency in the Office Setting. JAMA Intern Med. 2016;176(7):981-8.
261. Vilela-Martin JF, Vaz-de-Melo RO, Kuniyoshi CH, Neder A, Abdo R, Yugar-Toledo JC. Hypertensive crisis: clinical-epidemiological profile. Hypertens Res. 2011;34(3):367-71.
262. McNaughton CD, Self WH, Zhu YW, Janke AT, Storrow AB, Levy P. Incidence of Hypertension-Related Emergency Department Visits in the United States, 2006 to 2012. Am J Cardiol. 2015;116(11):1717-23.
263. Mancia G, Fagard R, Narkiewicz K, Redon J, Zanchetti A, Bohm M, et al. 2013 ESH/ESC Guidelines for themanagement of arterial hypertension The Task Force for the management ofarterial hypertension of the European Society ofHypertension (ESH) and of the European Society of Cardiology (ESC). J Hypertens. 2013;31(7):1281-357.
264. van den Born BJH, Beutler JJ, Gaillard CAJM, de Gooijer A, van den Meiracker AH, Kroon AA. Dutch guideline for the management of hypertensive crisis-2010 revision. Neth J Med. 2011;69(5):248-55.
265. Pinna G, Pascale C, Fornengo P, Arras S, Piras C, Panzarasa P, et al. Hospital Admissions for Hypertensive Crisis in the Emergency Departments: A Large Multicenter Italian Study. Plos One. 2014;9(4).
266. Bender SR, Fong MW, Heitz S, Bisognano JD. Characteristics and management of patients presenting to the emergency department with hypertensive urgency. J Clin Hypertens (Greenwich). 2006;8(1):12-8.
267. Leiba A, Cohen-Arazi O, Mendel L, Holtzman EJ, Grossman E. Incidence, aetiology and mortality secondary to hypertensive emergencies in a large-scale referral centre in Israel (1991-2010). J Hum Hypertens. 2016;30(8):498-502.
268. Chowdhury R, Khan H, Heydon E, Shroufi A, Fahimi S, Moore C, et al. Adherence to cardiovascular therapy: a meta-analysis of prevalence and clinical consequences. Eur Heart J. 2013;34(38):2940-8.
269. Herttua K, Tabak AG, Martikainen P, Vahtera J, Kivimaki M. Adherence to antihypertensive therapy prior to the first presentation of stroke in hypertensive adults: population-based study. Eur Heart J. 2013;34(38):2933-9.
270. van der Nagel BCH, Versmissen J, Bahmany S, van Gelder T, Koch BCP. High-throughput quantification of 8 antihypertensive drugs and active metabolites in human plasma using UPLC-MS/MS. J Chromatogr B Analyt Technol Biomed Life Sci. 2017;1060:367-73.
271. Sokolow M, Lyon TP. The ventricular complex in right ventricular hypertrophy as obtained by unipolar precordial and limb leads. Am Heart J. 1949;38(2):273-94.
272. Martin JF, Higashiama E, Garcia E, Luizon MR, Cipullo JP. Hypertensive crisis profile. Prevalence and clinical presentation. Arq Bras Cardiol. 2004;83(2):131-6; 25-30.
273. Al-Bannay R, Husain AA. Hypertensive crisis. Clinical presentation, comorbidities, and target organ involvement. Saudi Med J. 2010;31(8):916-20.

274. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. N Engl J Med. 2015;372(1):11-20.
275. Olsen MH, Angell SY, Asma S, Boutouyrie P, Burger D, Chirinos JA, et al. A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the Lancet Commission on hypertension. Lancet. 2016;388(10060):2665-712.
276. Salkic S, Batic-Mujanovic O, Ljuca F, Brkic S. Clinical presentation of hypertensive crises in emergency medical services. Mater Sociomed. 2014;26(1):12-6.
277. Merlo C, Bally K, Tschudi P, Martina B, Zeller A. Management and outcome of severely elevated blood pressure in primary care: a prospective observational study. Swiss Med Wkly. 2012;142:w13507.
278. McNaughton CD, Brown NJ, Rothman RL, Liu D, Kabagambe EK, Levy PD, et al. Systolic Blood Pressure and Biochemical Assessment of Adherence: A Cross-Sectional Analysis in the Emergency Department. Hypertension. 2017;70(2):307-14.
279. Dias E, Hachey B, McNaughton C, Nian H, Yu C, Straka B, et al. An LC-MS assay for the screening of cardiovascular medications in human samples. J Chromatogr B Analyt Technol Biomed Life Sci. 2013;937:44-53.
280. Saguner AM, Dur S, Perrig M, Schiemann U, Stuck AE, Burgi U, et al. Risk factors promoting hypertensive crises: evidence from a longitudinal study. Am J Hypertens. 2010;23(7):775-80.
281. Burnier M. Drug adherence in hypertension. Pharmacol Res. 2017;125(Pt B):142-9.
282. De Geest S, Ruppar T, Berben L, Schonfeld S, Hill MN. Medication non-adherence as a critical factor in the management of presumed resistant hypertension: a narrative review. EuroIntervention. 2014;9(9):1102-9.
283. de Souza WA, Sabha M, de Faveri Favero F, Bergsten-Mendes G, Yugar-Toledo JC, Moreno H. Intensive monitoring of adherence to treatment helps to identify "true" resistant hypertension. J Clin Hypertens (Greenwich). 2009;11(4):183-91.
284. Eskas PA, Heimark S, Eek Mariampillai J, Larstorp AC, Fadl Elmula FE, Hoieggen A. Adherence to medication and drug monitoring in apparent treatment-resistant hypertension. Blood Press. 2016;25(4):199-205.
285. Florczak E, Tokarczyk B, Warchol-Celinska E, Szwench-Pietrasz E, Prejbisz A, Gosk M, et al. Assessment of adherence to treatment in patients with resistant hypertension using toxicological serum analysis. A subgroup evaluation of the RESIST-POL study. Pol Arch Med Wewn. 2015;125(1-2):65-72.
286. Jung O, Gechter JL, Wunder C, Paulke A, Bartel C, Geiger H, et al. Resistant hypertension? Assessment of adherence by toxicological urine analysis. J Hypertens. 2013;31(4):766-74.
287. Pucci M, Martin U. Detecting non-adherence by urine analysis in patients with uncontrolled hypertension: rates, reasons and reactions. J Hum Hypertens. 2017;31(4):253-7.
288. Tomaszewski M, White C, Patel P, Masca N, Damani R, Hepworth J, et al. High rates of non-adherence to antihypertensive treatment revealed by high-performance liquid chromatography-tandem mass spectrometry (HP LC-MS/MS) urine analysis. Heart. 2014;100(11):855-61.
289. Brinker S, Pandey A, Ayers C, Price A, Raheja P, Arbique D, et al. Therapeutic drug monitoring facilitates blood pressure control in resistant hypertension. J Am Coll Cardiol. 2014;63(8):834-5.
290. Gupta P, Patel P, Strauch B, Lai FY, Akbarov A, Gulsin GS, et al. Biochemical Screening for Nonadherence Is Associated With Blood Pressure Reduction and Improvement in Adherence. Hypertension. 2017;70(5):1042-8.
291. Claxton AJ, Cramer J, Pierce C. A systematic review of the associations between dose regimens and medication compliance. Clin Ther. 2001;23(8):1296-310.

292. Kamran A, Sadeghieh Ahari S, Biria M, Malepour A, Heydari H. Determinants of Patient's Adherence to Hypertension Medications: Application of Health Belief Model Among Rural Patients. Ann Med Health Sci Res. 2014;4(6):922-7.
293. Williams B, MacDonald TM, Morant S, Webb DJ, Sever P, McInnes G, et al. Spironolactone versus placebo, bisoprolol, and doxazosin to determine the optimal treatment for drug-resistant hypertension (PATHWAY-2): a randomised, double-blind, crossover trial. Lancet. 2015;386(10008):2059-68.
294. Zeller A, Taegtmeyer A, Martina B, Battegay E, Tschudi P. Physicians' ability to predict patients' adherence to antihypertensive medication in primary care. Hypertens Res. 2008;31(9):1765-71.
295. Kronish IM, Woodward M, Sergie Z, Ogedegbe G, Falzon L, Mann DM. Meta-analysis: impact of drug class on adherence to antihypertensives. Circulation. 2011;123(15):1611-21.
296. Nielen M, Poos M, Gommer A. Chronische aandoeningen en multimorbiditeit 2019 [Available from: https://www.volksgezondheidenzorg.info/onderwerp/chronische-ziekten-en-multimorbiditeit/cijfers-context/huidige-situatie#node-prevalentie-multimorbiditeit-naar-aantal-chronische-aandoeningen.
297. Gosling C, Gorzeman M, Offeringa-Klooster M, Berdowski J. Brandbrief 'Regionale spoedzorg - de rek is er uit'2016. Available from: https://www.spoedzorgnet.nl/sites/default/files/documents/brandbrief_regionale_spoedzorg_-_de_rek_is_er_uit_2016-05-19.pdf.
298. Van Der Linden MC, Khursheed M, Hooda K, Pines JM, Van Der Linden N. Two emergency departments, 6000 km apart: Differences in patient flow and staff perceptions about crowding. Int Emerg Nurs. 2017;35:30-6.
299. Fitzpatrick D, Doyle K, Finn G, Gallagher P. The collateral history: an overlooked core clinical skill. Eur Geriatr Med. 2020;11(6):1003-7.
300. Emond M, Grenier D, Morin J, Eagles D, Boucher V, Le Sage N, et al. Emergency Department Stay Associated Delirium in Older Patients. Can Geriatr J. 2017;20(1):10-4.
301. Perdahl T, Axelsson S, Svensson P, Djarv T. Patient and organizational characteristics predict a long length of stay in the emergency department - a Swedish cohort study. Eur J Emerg Med. 2017;24(4):284-9.
302. Bonin-Guillaume S, Durand AC, Yahi F, Curiel-Berruyer M, Lacroix O, Cretel E, et al. Predictive factors for early unplanned rehospitalization of older adults after an ED visit: role of the caregiver burden. Aging Clin Exp Res. 2015;27(6):883-91.
303. Horney C, Schmader K, Sanders LL, Heflin M, Ragsdale L, McConnell E, et al. Health care utilization before and after an outpatient ED visit in older people. Am J Emerg Med. 2012;30(1):135-42.
304. Nagurney JM, Fleischman W, Han L, Leo-Summers L, Allore HG, Gill TM. Emergency Department Visits Without Hospitalization Are Associated With Functional Decline in Older Persons. Ann Emerg Med. 2017;69(4):426-33.
305. Ayaz T, Sahin SB, Sahin OZ, Bilir O, Rakici H. Factors affecting mortality in elderly patients hospitalized for nonmalignant reasons. J Aging Res. 2014;2014:584315.
306. Wu LW, Chen WL, Peng TC, Chiang ST, Yang HF, Sun YS, et al. All-cause mortality risk in elderly individuals with disabilities: a retrospective observational study. BMJ Open. 2016;6(9):e011164.
307. Alam N, Hobbelink EL, van Tienhoven AJ, 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.
308. Panday RSN, Minderhoud TC, Alam N, Nanayakkara PWB. Prognostic value of early warning scores in the emergency department (ED) and acute medical unit (AMU): A narrative review. European Journal of Internal Medicine. 2017;45:20-31.

309. Spencer W, Smith J, Date P, de Tonnerre E, Taylor DM. Determination of the best early warning scores to predict clinical outcomes of patients in the emergency department. Emerg Med J. 2019;36(12):716-21.
310. Holdstock V, Shaw M, McPeake J, Kinsella J. Validation of Qsofa and Comparison with Sirs, News, and Sofa in a Glasgow Patient Cohort. Critical Care Medicine. 2018;46(1):733-.
311. Usman OA, Usman AA, Ward MA. Comparison of SIRS, qSOFA, and NEWS for the early identification of sepsis in the Emergency Department. Am J Emerg Med. 2019;37(8):1490-7.
312. Guarino M, Gambuti E, Alfano F, De Giorgi A, Maietti E, Strada A, et al. Predicting in-hospital mortality for sepsis: a comparison between qSOFA and modified qSOFA in a 2-year single-centre retrospective analysis. Eur J Clin Microbiol Infect Dis. 2020.
313. Liu S, He C, He W, Jiang T. Lactate-enhanced-qSOFA (LqSOFA) score is superior to the other four rapid scoring tools in predicting in-hospital mortality rate of the sepsis patients. Ann Transl Med. 2020;8(16):1013.
314. Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 Update. Crit Care Med. 2018;46(6):997-1000.
315. Gensini GF, Conti AA. The evolution of the concept of 'fever' in the history of medicine: from pathological picture per se to clinical epiphenomenon (and vice versa). J Infect. 2004;49(2):85-7.
316. Walter EJ, Hanna-Jumma S, Carraretto M, Forni L. The pathophysiological basis and consequences of fever. Crit Care. 2016;20(1):200.
317. Matsuno O, Kataoka H, Takenaka R, Okubo F, Okamoto K, Masutomo K, et al. Influence of age on symptoms and laboratory findings at presentation in patients with influenza-associated pneumonia. Arch Gerontol Geriat. 2009;49(2):322-5.
318. Metlay JP, Schulz R, Li YH, Singer DE, Marrie TJ, Coley CM, et al. Influence of age on symptoms at presentation in patients with community-acquired pneumonia. Archives of Internal Medicine. 1997;157(13):1453-9.
319. Norman DC. Fever in the elderly. Clin Infect Dis. 2000;31(1):148-51.
320. Honaker T, Long A, Lyons D, McCoy R, Salters D, MacArthur SK. Monitoring Temperature Knowledge and skills of outpatients with cancer. Clin J Oncol Nurs. 2018;22(6):619-25.
321. Patel DM, Riedel DJ. Fever in Immunocompromised Hosts. Emerg Med Clin N Am. 2013;31(4):1059-+.
322. Del Bene VE. Temperature. 1990.
323. Fumagalli R, Bellani G, Perri A. Which drugs for the control of fever in critical patients. Curr Drug Targets. 2009;10(9):881-6.
324. Blomaard LC, Speksnijder C, Lucke JA, de Gelder J, Anten S, Schuit SCE, et al. Geriatric Screening, Triage Urgency, and 30-Day Mortality in Older Emergency Department Patients. J Am Geriatr Soc. 2020;68(8):1755-62.
325. Brabrand M, Kellett J. Mobility measures should be added to the National Early Warning Score (NEWS). Resuscitation. 2014;85(9):e151.
326. Brabrand M, Kellett J, Opio M, Cooksley T, Nickel CH. Should impaired mobility on presentation be a vital sign? Acta Anaesthesiol Scand. 2018;62(7):945-52.
327. Nickel CH, Kellett J, Nieves Ortega R, Lyngholm L, Wasingya-Kasereka L, Brabrand M. Mobility Identifies Acutely Ill Patients at Low Risk of In-Hospital Mortality: A Prospective Multicenter Study. Chest. 2019;156(2):316-22.
328. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142-8.

329. Oud FM, de Rooij SE, Schuurman T, Duijvelaar KM, van Munster BC. [Predictive value of the VMS theme 'Frail elderly': delirium, falling and mortality in elderly hospital patients] Voorspellende waarde VMS-thema 'Kwetsbare ouderen': delirium, valincidenten en overlijden bij oudere ziekenhuispatienten. Ned Tijdschr Geneeskd. 2015;159:A8491.
330. Schuijt HJ, Oud FMM, Bruns EJR, van Duijvendijk P, Van der Zaag-Loonen HJ, Spies PE, et al. Does the Dutch Safety Management Program predict adverse outcomes for older patients in the emergency department? Neth J Med. 2020;78(5):244-50.
331. Goedecke T, Ord K, Newbould V, Brosch S, Arlett P. Medication Errors: New EU Good Practice Guide on Risk Minimisation and Error Prevention. Drug Saf. 2016;39(6):491-500.
332. Barber N. Should we consider non-compliance a medical error? Qual Saf Health Care. 2002;11(1):81-4.
333. Warle-van Herwaarden MF, Valkhoff VE, Herings RM, Engelkes M, van Blijderveen JC, Rodenburg EM, et al. Quick assessment of drug-related admissions over time (QUADRAT study). Pharmacoepidemiol Drug Saf. 2015;24(5):495-503.
334. Leendertse AJ, Egberts AC, Stoker LJ, van den Bemt PM, Group HS. Frequency of and risk factors for preventable medication-related hospital admissions in the Netherlands. Arch Intern Med. 2008;168(17):1890-6.
335. Mitchell Scott B, Considine J, Botti M. Medication errors in ED: Do patient characteristics and the environment influence the nature and frequency of medication errors? Australas Emerg Nurs J. 2014;17(4):167-75.
336. Fleischmann C, Scherag A, Adhikari NK, Hartog CS, Tsaganos T, Schlattmann P, et al. Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations. Am J Respir Crit Care Med. 2016;193(3):259-72.
337. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med. 2006;34(6):1589-96.
338. De Waele JJ, Dhaese S. Antibiotic stewardship in sepsis management: toward a balanced use of antibiotics for the severely ill patient. Expert Rev Anti-Infe. 2019;17(2):89-97.
339. WHO. WHO Global Strategy for Containment of Antimicrobial Resistance. Switzerland2001.
340. Bassetti M, Rello J, Blasi F, Goossens H, Sotgiu G, Tavoschi L, et al. Systematic review of the impact of appropriate versus inappropriate initial antibiotic therapy on outcomes of patients with severe bacterial infections. Int J Antimicrob Agents. 2020;56(6):106184.
341. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58(5):377-82.
342. Lamy B, Dargere S, Arendrup MC, Parienti JJ, Tattevin P. How to Optimize the Use of Blood Cultures for the Diagnosis of Bloodstream Infections? A State-of-the Art. Front Microbiol. 2016;7.
343. Trung NT, Thau NS, Bang MH, Song LH. PCR-based Sepsis@Quick test is superior in comparison with blood culture for identification of sepsis-causative pathogens. Sci Rep. 2019;9(1):13663.
344. Cendejas-Bueno E, Romero-Gomez MP, Mingorance J. The challenge of molecular diagnosis of bloodstream infections. World J Microbiol Biotechnol. 2019;35(4):65.
345. Skvarc M, Stubljar D, Rogina P, Kaasch AJ. Non-culture-based methods to diagnose bloodstream infection: Does it work? Eur J Microbiol Immunol (Bp). 2013;3(2):97-104.
346. van der Does Y, Limper M, Jie KE, Schuit SCE, Jansen H, Pernot N, et al. Procalcitonin-guided antibiotic therapy in patients with fever in a general emergency department population: a multicentre non-inferiority randomized clinical trial (HiTEMP study). Clin Microbiol Infect. 2018;24(12):1282-9.

347. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Community Health. 2020;8(1):e000262.
348. : Oxford University Press 2001. Paperback Oxford English Dictionary
349. Portet S. A primer on model selection using the Akaike Information Criterion. Infect Dis Model. 2020;5:111-28.
350. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol. 2013;64(5):402-6.
351. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj-Brit Med J. 2009;339.
352. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol. 2003;56(9):826-32.
353. Nieboer D, Vergouwe Y, Ankerst DP, Roobol MJ, Steyerberg EW. Improving prediction models with new markers: a comparison of updating strategies. BMC Med Res Methodol. 2016;16(1):128.
354. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol. 2006;6:54.
355. Deo RC. Machine Learning in Medicine. Circulation. 2015;132(20):1920-30.
356. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLoS One. 2018;13(7):e0201016.
357. Gravesteijn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, van Calster B, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol. 2020;122:95-107.
358. Chaisirin W, Wongkrajang P, Thoesam T, Praphruetkit N, Nakornchai T, Riyapan S, et al. Role of Point-of-Care Testing in Reducing Time to Treatment Decision-Making in Urgency Patients: A Randomized Controlled Trial. West J Emerg Med. 2020;21(2):404-10.
359. Thamburaj R, Sivitz A. Does the use of bedside pelvic ultrasound decrease length of stay in the emergency department? Pediatr Emerg Care. 2013;29(1):67-70.
360. Wilson SP, Connolly K, Lahham S, Subeh M, Fischetti C, Chiem A, et al. Point-of-care ultrasound versus radiology department pelvic ultrasound on emergency department length of stay. World J Emerg Med. 2016;7(3):178-82.
361. Chrusciel J, Fontaine X, Devillard A, Cordonnier A, Kanagaratnam L, Laplanche D, et al. Impact of the implementation of a fast-track on emergency department length of stay and quality of care indicators in the Champagne-Ardenne region: a before-after study. BMJ Open. 2019;9(6):e026200.
362. Kwa P, Blake D. Fast track: Has it changed patient care in the emergency department? Emerg Med Australas. 2008;20(1):10-5.
363. Rijksoverheid. Coronavirus dashboard: Rijksoverheid; 2020 [Available from: https://coronadashboard.government.nl/landelijk/ziekenhuis-opnames.
364. COVID: Interne Geneeskunde Rotterdam; 2020 [Available from: https://www.internegeneeskunderotterdam.nl/covid/.

Appendices

Authors and Affiliations

Authors in alphabetical order. Affiliations at the time that research was conducted.

Jelmer Alsma, MD, PhD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Sander Anten, MD
Department of Internal Medicine, Alrijne Hospital, Leiderdorp, the Netherlands

Lodewijk A.A.M. van Attekum, MD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands

Soma Bahmany, BSc
Department of Hospital Pharmacy , Erasmus University Medical Center, Rotterdam, The Netherlands

Laura C. Blomaard, MD, PhD
Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

Wichor M. Bramer, PhD
Medical Library, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands

Anniek Brink, MD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Hans S. Brink, MD, PhD
Department of Internal Medicine, Medisch Spectrum Twente, Enschede, The Netherlands

Willian van Dijk, MD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands

Aniska W. Fortuin, MD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands

Jelle de Gelder, MD, PhD
Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

Bas de Groot, MD, PhD
Department of Emergency Medicine, Leiden University Medical Center, Leiden, The Netherlands

Edon Hameli, MD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Birgit C.P. Koch, PhD
Department of Hospital Pharmacy , Erasmus University Medical Center, Rotterdam, The Netherlands

Hester F. Lingsma, PhD
Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands

Jacinta A. Lucke, MD, PhD
Department of Emergency Medicine, Leiden University Medical Center, Leiden, The Netherlands

Anton H. van den Meiracker, MD, PhD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Damian C. Melles, PhD
Department of Medical Microbiology and Infectious Diseases, Erasmus University Medical Center, Rotterdam, the Netherlands

Simon P. Mooijaart, MD, PhD
Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

Naomi Overgaauw, MD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands

Laura E.J. Peeters, MSc
Department of Internal Medicine and Hospital Pharmacy , Erasmus University Medical Center, Rotterdam, The Netherlands

Pleunie P.M. Rood, MD, PhD
Department of Emergency Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Stephanie C.E. Schuit, MD, PhD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Romy Schuttevaer, MD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Jurriaan E.M. de Steenwinkel, PhD
Department of Medical Microbiology and Infectious Diseases, Erasmus University Medical Center, Rotterdam, The Netherlands

Annelies Verbon, MD, PhD
Department of Internal Medicine and Medical Microbiology and Infectious Diseases, Erasmus University Medical Center, Rotterdam, the Netherlands

Rob J.C.G. Verdonschot, MD, PhD
Department of Emergency Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Jorie Versmissen, MD, PhD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Robert Zietse, MD, PhD
Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands

Appendices

Abbreviations

0-♀

°C	Degrees Celsius
♀	Female
♂	Male
6-CIT	Six Item Cognitive Impairment Test

α-Ω

α	Calibration Intercept
β	Calibration Slope
µL	Microlitre
X^2	Chi-squared

A

aAISS	adjusted Acute Illness Severity Score
AB	Anniek Brink
ABCDE-method	Airway Breathing Circulation Disability Environment-method
ACE	Angiotensin Converting Enzyme
ADM	Admitted
AEAT	Appropriate Empiric Antibiotic Therapy
AF	Aniska Fortuin
AIC	Akaike Information Criterion
AISS	Acute Illness Severity Score
AM	Ante Meridiem
AMU	Acute Medical Unit
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
APACHE II	Acute Physiologic Assessment and Chronic Health Evaluation II
APOP	Acutely Presenting Older Patient
AUC	Area Under the Curve
AVPU	Alert Verbal Pain Unresponsive

B

BP	Blood Pressure
bpm	Beats per Minute
BS	Brier Score
BSI	Bloodstream Infection

C

CAM-ICU	Confusion Assessment Method for the Intensive Care Unit
CCI	Charlson comorbidity index

CENTRAL	Cochrane Central Register of Controlled Trials
CHARMS	Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies
CI	Confidence Interval
CLEARED	CalcuLation of the Elderly Admission Risk in the Emergency Department
COVID-19	Coronavirus Disease 2019
CRP	C-Reactive Protein
CS	Calibration Slope
CTR	Cardiothoracic ratio
CURB-65	Confusion, blood Urea nitrogen, Respiratory rate, Blood pressure, age 65

D

DBP	Diastolic Blood Pressure
DC	Discharged
dL	Decilitre
DV	Derivation

E

E. coli	Escherichia Coli
e.g.	Exempli Gratia
ECG	Electrocardiogram
ED	Emergency Department
ED LOS	Length of Stay at the Emergency Department
EPICS	Emergency Processes In Clinical Structures
Erasmus MC	Erasmus University Medical Center
ESI	Emergency Severity Index
et al.	Et Alii
EWS	Early Warning Score

F

FiO2	Fraction of Inspired Oxygen

G

g	Gram
GAPS	Glasgow Admission Prediction Score
GCS	Glasgow Coma Scale
GOF	Goodness of Fit
GP	General Practitioner

H

HE	Hypertensive Emergency

hiTEMP	Higher diagnostic accuracy and cost effectiveness using procalcitonin in the Treatment of Emergency Medicine Patients with fever
HL GOF	Hosmer-Lemeshow Goodness of Fit
HMC Bronovo	Haaglanden Medical Center, location Bronovo
HOTEL	Hypotension, Oxygen saturation, low Temperature, ECG changes and Loss of independence
HU	Hypertensive Urgency

I

i.e.	Id Est
ICU	Intensive Care Unit
IP-10	Interferon Gamma-induced Protein 10
IQR	Interquartile Range
ISAR	Identification of Seniors At Risk

J

JA	Jelmer Alsma

K

Katz-ADL	Katz Index of Independence in Activities of Daily Living

L

L	Litre
LA	Lodewijk Anna Antonius Maria van Attekum
LC-MS/MS	Liquid Chromatography- Tandem Mass Spectrometry
LOS	Length of Stay
LR	Likelihood Ratio
LRM	Logistic Regression Model
LRχ2	Likelihood Ratio Chi-squared
LUMC	Leiden University Medical Center
LVH	Left Ventricular Hypertrophy

M

MAP	Mean Arterial Pressure
MARS	Medical Admissions Risk System
MC	Multicentre
MEOWS	Modified Early Obstetric Warning System
MEWS	Modified Early Warning Score
mg	Milligram
min	Minute
MIU	Minor Injury Unit
mL	Millilitre

mmHg	Millimetre of Mercury
mmol	Millimole
MST	Medisch Spectrum Twente
MTS	Manchester Triage System

N

n	Number
NEWS	National Early Warning Score
NHS	United Kingdom National Health Service
NPV	Negative Predictive Value
NS	Not Specified
NSAID	NonSteroidal Anti-Inflammatory Drugs
NTS	Netherlands Triage System

O

OR	Odds Ratio

P

p	Probability
PaCO2	Partial Pressure of Carbon Dioxide
PARIS	Pressure, Age, Respiratory rate, loss of Independence and peripheral oxygen Saturation.
PCR	Polymerase Chain Reaction
PCSS	Prospective Cross-Sectional Study
PEWS	Pediatric Early Warning Score
PhD	Doctor of Philosophy
PIRO	Predisposition, Infection, Response and Organ Failure
PM	Post Meridiem
POCS	Prospective Observational Cohort Study
POCT	Point-of-Care Testing
POCUS	Point-of-Care UltraSound
PPV	Positive Predictive Value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	International Prospective Register of Systematic Reviews
PSS	Physiological Scoring System
p-value	Probability value

Q

qSOFA	quick Sepsis-related Organ Failure Assessment

R

RAPS	Rapid Acuity Physiology Score

RBC	Red Blood Cell Count
RCSS	Retrospective Cross-Sectional Study
RDW	Red Cell Distribution Width
Ref	Reference Category
REMS	Rapid Emergency Medicine Score
ROC	Receiver Operating Characteristic
ROCS	Retrospective Observational Cohort Study
RR	Respiratory Rate
RS	Romy Schuttevaer

S

S. Epidermidis	Staphylococcus Epidermidis
SaO2	Pulse Oximetry Saturation
SBP	Systolic Blood Pressure
SC	Single Centre
SCS	Simple Clinical Score
SD	Standard Deviation
SE	Standard Error
Sepsis-3	Third International Consensus Definition for Sepsis
SIRS	Systemic Inflammatory Response Syndrome
SOFA	Sequential Organ Failure Assessment score
SpO2	Peripheral Oxygen Saturation
SPSS	Statistical Package for the Social Science

T

t	Timepoint
TNP	Triage Nurse Prediction
TRAIL	Tumor Necrosis Factor-related Apoptosis-Inducing Ligand
TRST	Triage Risk Screening Tool

U

UK	United Kingdom
USA	United States of America

V

VAS	Visual Analogue Scale
VD	Validation
VMS	Dutch Safety Management Program
vs.	Versus
VSS	Vital Sign Score

W

WBC	White Blood Cell Count
WMB	Wichor Matthijs Bramer
WMO	Medical Research Involving Human Subjects Act

Y

y	Year

Appendices

PhD Portfolio

Summary of PhD training and teaching activities

PhD student:	Anniek Brink
Erasmus MC Department:	Department of Internal Medicine, Section Acute Medicine
Research School:	Netherlands Institute for Health Sciences (NIHES)
PhD period:	2016-2020
Promotors:	prof.dr.S.C.E. Schuit, prof.dr. R. Zietse, prof.dr. H.F. Lingsma
Co-promotors:	dr. J. Alsma
Date of thesis defence:	9th of February 2022

	Period	ECTS
1. PhD training		
General courses		
BROK course	2017	1.5
Scientific Writing in English for Publication, NIHES	2016	2.0
Scientific Integrity	2019	0.3
Specific courses		
Study design, NIHES	2014	4.3
Biostatistical methods I: basic principles, NIHES	2014	5.7
Biostatistical methods II: classical regression models, NIHES	2014	4.3
Clinical Epidemiology, NIHES	2014	5.7
Methodologic Topics in Epidemiologic Research, NIHES	2014	1.4
Methods of Public Health Research, NIHES	2015	0.7
Fundamentals of Medical Decision Making, NIHES	2015	0.7
Advanced Analysis of Prognosis Studies, NIHES	2016	0.9
Principles of Epidemiologic Data-analysis, NIHES	2016	0.7
Advanced Topics in Decision-making in Medicine, NIHES	2015	1.9
Oral presentations		
Erasmus MC Honours Alumni Vereniging, Rotterdam	2015	0.7
Science days Internal Medicine, Antwerp, Belgium	2016	0.7
Dutch Acute Medicine Conference, Enschede	2016	0.7
Dutch Emergency Medicine Conference, Egmond aan Zee	2017	0.7
Dutch Acute Medicine Conference, Groningen	2017	0.7
Internal Medicine Research Meeting, Rotterdam	2017	0.7
Nederlandse Vereniging van Interne Acute Geneeskunde conference, Rotterdam	2019	0.7
Science days Internal Medicine, St. Michielsgestel	2020	0.7
Poster presentations		
Science days Internal Medicine, Antwerp, Belgium	2017	0.3
Social Media and Critical Care Conference, Berlin, Germany	2017	0.3

	Period	ECTS
Abstracts		
Internist days, Maastricht	2017	
(Inter)national conferences		
Social Media and Critical Care Conference, Chicago, USA	2015	0.8
Dutch Emergency Medicine Conference, Egmond aan Zee	2014	0.8
Dutch Emergency Medicine Conference, Egmond aan Zee	2015	0.8
European Society of Emergency Medicine (EuSEM), Amsterdam	2014	0.8
2. Teaching		
(Co-)supervising master thesis (4x)		
Naomi Overgaauw. Project title: Drug non-adherence as major cause of severe hypertension at the emergency department	2016	1.5
Willian van Dijk. Project title: Appropriate empirical treatment of bacteremia and adherence to guidelines at a Dutch Emergency Department.	2017	1.5
Aniska Fortuin. Project title: The identification of prediction models for mortality in adult patients visiting the Emergency Department: a systematic review	2017	1.5
Lodewijk van Attekum. Project title: The identification of prediction models on admission and length of stay in adults at the Emergency Department: a systematic review	2017	1.5
Teaching		
Lecture NIHES Clinical epidemiology course	2017	0.7
3. Organizational skills		
Chairing the acute medicine research meetings	2016-'17	0.5
4. Grants and prizes		
Best oral presentation, DAM conference, Enschede	2016	
Erasmus MC Trustfonds Travel grant	2017	
SMACC bursary	2015	
SMACC bursary	2017	
5. Other		
Reviewing activities for several peer-reviewed journals including Peer journal, Plos One, Internal and Emergency Medicine, Netherlands Journal of Medicine, European Journal of Internal Medicine	2016-'21	2
Total workload (ECTS)		**48.7**

Appendices

Bibliography

Manuscripts within this thesis

Predicting inhospital admission at the emergency department: a systematic review.
Brink A, Alsma J, van Attekum LAAM, Bramer WM, Zietse R, Lingsma HF, Schuit SCE.
Emergency Medicine Journal. 2021 Oct 28;emermed-2020-210902.

Prediction admission in the older population in the Emergency Department: the CLEARED tool.
Brink A, Alsma J, Brink HS, de Gelder J, Lucke JA, Mooijaart SP, Zietse R, Schuit SCE, Lingsma HF. *Netherlands Journal of Medicine. 2020 Dec;78(6):357-367.*

Prediction models for mortality in adult patients visiting the Emergency Department: a systematic review.
Brink A, Alsma J, Fortuin AW, Bramer WM, Zietse R, Lingsma HF, Schuit SCE.
Acute Medicine. 2019;18(3):171-183.

Predicting mortality in patients with suspected sepsis at the Emergency Department; a retrospective cohort study comparing qSOFA, SIRS and National Early Warning Score.
Brink A*, Alsma J*, Verdonschot RJCG, Rood PPM, Zietse R, Lingsma HF, Schuit SCE.
PLoS One. 2019;14(1):e0211133.

Predicting 30-day mortality using point-of-care testing; an external validation and derivation study.
Brink A, Schuttevaer R, Alsma J, Zietse R, Schuit SCE, Lingsma HF.
PLoS One. 2020 Sep 24;15(9):e0239318.

Non-adherence to antimicrobial guidelines in patients with bloodstream infection visiting the emergency department.
Schuttevaer R*, Brink A*, Alsma J, van Dijk W, Melles DC, de Steenwinkel JEM, Lingsma HF, Verbon A, Schuit SCE.
European Journal of Internal Medicine. 2020;78:69-75.

The association of body temperature with antibiotic therapy and mortality in patients attending the emergency department with suspected infection.
Schuttevaer R, Brink A, Alsma J, de Steenwinkel JEM, Verbon A, Schuit SCE, Lingsma HF.
European Journal of Emergency Medicine. 2021 Dec 1;28(6):440-447.

Appropriate Empirical Antibiotic Therapy and Mortality: Conflicting Data Explained by Residual Confounding.
Schuttevaer R*, Alsma J*, Brink A, van Dijk W, de Steenwinkel JEM, Lingsma HF, Melles DC, Schuit SCE.
PLoS One. 2019;14(11):e0225478.

Drug nonadherence is a common but often overlooked cause of hypertensive urgency and emergency at the emergency department.
Overgaauw N*, Alsma J*, Brink A, Hameli E, Bahmany S, Peeters LEJ, Van Den Meiracker AH, Schuit SCE, Koch BCP, Versmissen J.
Journal of Hypertension. 2019 May;37(5):1048-1057.

Other publications

CAM-ICU may not be the optimal screening tool for early delirium screening in older emergency department patients: a prospective cohort study.
Lucke JA, De Gelder J, Blomaard LC, Fogteloo AJ, Alsma J, Schuit SCE, Brink A, De Groot B, Blauw GJ, Mooijaart SP.
European Journal of Emergency Medicine. 2019 Dec;26(6):428-432.

Optimization of the APOP screener to predict functional decline or mortality in older emergency department patients: Cross-validation in four prospective cohorts.
De Gelder J, Lucke JA, Blomaard LC, Booijen AM, Fogteloo AJ, Anten S, Steyerberg EW, Alsma J, Schuit SCE, Brink A, De Groot B, Blauw GJ, Mooijaart SP.
Experimental Gerontology. 2018;110:253-9.

Vital signs and impaired cognition in older emergency department patients: The APOP study.
Lucke JA, de Gelder J, Blomaard LC, Heringhaus C, Alsma J, Schuit SCE, Brink A, Anten S, Blauw GJ, de Groot B, Mooijaart SP.
PLoS One. 2019 Jun 20;14(6):e0218596.

Palliatieve sedatie voor covid-19-patiënten kan veel beter; Ikazia Ziekenhuis past protocol aan. ten Cate M, Brink A, Struik EM, Thijs-Visser MF.
Medisch Contact. 2021 Jul 29:18-20.

*These authors contributed equally to this work.

Appendices

Dankwoord

Dit proefschrift is voortgekomen uit wetenschappelijk onderzoek verricht op de afdeling Inwendige Geneeskunde en de Spoedeisende Hulp van het Erasmus Medisch Centrum te Rotterdam. Het dankwoord is misschien wel het leukste hoofdstuk van het proefschrift om te schrijven (en te lezen). Een promotieonderzoek doe je niet alleen en ik wil dan ook iedereen bedanken die me de afgelopen jaren heeft bijgestaan op professioneel, maar ook op persoonlijk en emotioneel vlak. Zonder jullie steun en support zou dit proefschrift niet tot stand zijn gekomen. Zonder iemand tekort te willen doen, wil ik een aantal personen in het bijzonder bedanken.

Gedurende mijn tijd als promovendus is er enigszins geschoven in mijn promotiecommissie. Prof.dr. Stephanie C.E. Klein Nagelvoort-Schuit werd als tweede promotor aangesteld, die op haar beurt de rol als een van de copromotoren overdroeg aan dr. Jelmer Alsma. De meest recente verandering is dat mijn tweede copromotor, prof.dr. Hester F. Lingsma, mijn derde promotor is geworden.

Allereerst mijn eerste promotor, **prof.dr. Bob Zietse**. Beste Bob, in de eerste weken van de studie Geneeskunde stond ik al bij u op de stoep om onderzoek te kunnen doen. Hoewel ik van u eerst de resultaten moest afwachten van de eerste tentamens, kon ik al snel aansluiten bij een onderzoek. Ik begon met het invoeren van data bij een onderzoek naar de anion-gap en ik vind het wel een beetje spijtig dat dit hoofdstuk het uiteindelijk niet heeft gered om toegevoegd te worden aan dit proefschrift. Ik waardeer uw begeleiding en wil u bedanken voor de tijd en energie die u in me heeft geïnvesteerd om me te laten ontwikkelen als wetenschapper. Ook heb ik enorm respect voor uw enthousiasme voor uw vak en de manier waarop u jonge collega's met de nodige (droge) humor enthousiasmeert voor de nefrologie. Ondanks dat het onderwerp van dit proefschrift niet geheel in uw onderzoeksveld past, hebben we samen ervoor gezorgd dat de verschillende onderdelen samen in dit boekje terecht kwamen.

Prof.dr. Stephanie C.E. Klein Nagelvoort-Schuit Beste Stephanie, dankjewel voor het enorme vertrouwen. Zonder dit vertrouwen had ik geen database gehad met onderzoeksgegevens en had dit boekje niet in deze vorm bestaan. Ik vind het fantastisch dat je hoogleraar acute geneeskunde bent geworden en dat ik daardoor de luxe heb dat ik inmiddels drie promotoren heb. Het is een eer dat ik je eerste promovendus mag zijn, waarbij jij de rol van promotor hebt. Veel dank voor je aanstekelijke enthousiasme en oplossende vermogen. Door jouw enorme hoeveelheid connecties was elk probleem spoedig opgelost. Ik bewonder je werkwijze, altijd druk met van alles, maar toch de tijd nemen om ver buiten kantoortijden nog te reageren op mijn mails, manuscripten et cetera. Ooit hoop ik net zo'n alleskunner te zijn als jij!

Prof.dr. Hester F. Lingsma, Beste Hester, zonder jou had ik denk ik nog steeds in spagaat gezeten in de statistiek. Wat ben ik blij dat jij (op de valreep) mijn promotor bent geworden. Met jouw hulp heb ik de statistiek me steeds meer eigen kunnen maken. Ik vind het bewonderenswaardig dat je, ondanks dat je geen arts bent en niet thuis bent in de (acute) interne geneeskunde, je de stof gemakkelijk je eigen kan maken om vervolgens zeer nuttige opmerkingen met betrekking tot het onderzoek te kunnen geven. Erg bedankt voor je vertrouwen en begeleiding. Ik hoop dat we ook in de toekomst nog veelvuldig samen mogen werken. Gefeliciteerd met het bereiken van het hoogleraarschap en ik wens je veel succes in je carrière als kersverse professor!

Voor de totstandkoming van dit boekje is **dr. Jelmer Alsma** essentieel. Beste Jelmer, bij jou kwam ik terecht als onervaren student en zocht ik data op voor het anion-gap onderzoek. Je kan wel stellen dat het onderzoek doen behoorlijk uit de hand is gelopen. Samen verkenden we de (on)mogelijkheden van de database en dit heeft tot dusver geresulteerd in twee proefschriften en bijna drie. Jouw deur en Whatsapp stonden altijd open voor vragen, voor momenten om te discussiëren en nieuwe plannen te maken of over specifieke bewoordingen in een manuscript. Bedankt voor al je hulp en levenslessen. Ik hoop dat we ook na de verdediging met eenzelfde enthousiasme het onderzoek binnen de acute interne geneeskunde kunnen continueren. Ik kijk er naar uit om tijdens mijn opleidingstijd nog veel meer van je te leren in de kliniek.

Daarnaast wil ik de leden van de kleine commissie, **prof.dr. Jan van Saasse**, **prof.dr. Prabath Nanayakkara** en **prof.dr. Francesco Mattace-Raso** bedanken. Dank dat jullie de tijd en moeite hebben genomen om mijn proefschrift te beoordelen. Jan van Saasse, ik vind het zo leuk dat u al bij een aantal belangrijke mijlpalen in mijn medische carrière betrokken bent geweest. Eerst bij het uitreiken van mijn eerste doktersjas tijdens de Witte Jassen Ceremonie, vervolgens bij mijn arts-examen en nu als secretaris van de kleine commissie. Beste Prabath, ik kijk met plezier terug naar de ORCA meetings en NVIAG congressen. U bent een inspiratie voor me. Francesco, fijn dat u met uw kritische blik vanuit de sectie ouderengeneeskunde mijn proefschrift hebt bekeken, gezien een belangrijk deel van mijn proefschrift en de populatie patiënten op de SEH ouderen betreft.

Eveneens wil ik **prof.dr. Nathalie van der Velde, dr. Tjebbe Hagenaars** bedanken. Het is voor mij een groot genoegen dat jullie zitting willen nemen in de oppositie. Ik zie uit naar onze gedachtewisseling op 9 februari 2022. **dr. John Kellet,** thank you for your willingness to travel to Rotterdam to take place in the committee if the situation regarding covid-19 allows.

Ik wil ook alle coauteurs bedanken. Hartelijk dank voor jullie kritische en waardevolle feedback op de diverse manuscripten. In het bijzonder allereerst **Dr. Wichor Bramer**. Met uw nauwkeurige en zeer uitgebreide zoektermen ben ik in staat geweest om twee systematic reviews te schrijven. Het heeft een tijd en veel nieuwe zoek strategieën geduurd, maar ook het tweede review is eindelijk gepubliceerd!

Verder wil ik iedereen die betrokken is geweest bij de APOP-studie, maar met name: **Jelle de Gelder**, **Jacinta Lucke**, **Laura Blomaard** en **dr. S. Mooijaart** bedanken. Een deel van de hoofdstukken in dit boekje is gebaseerd op data die jullie hebben aangeleverd. Dankzij jullie heb ik met behulp van de ZonMW beurs een jaar betaald promotieonderzoek kunnen doen voor aanvang van de coschappen.

Ook wil ik Marcel Eitink bedanken voor het organiseren van een database van patiënten van de SEH van het MST voor het CLEARED artikel. Het was gezellig om twee weken in een hok opgesloten te zitten om de juiste data te verkrijgen.

De basis voor diverse hoofdstukken in dit boekje is gelegd door diverse masteronderzoekstudenten Geneeskunde. Daarom wil ik heel graag **Aniska**, **Willian**, **Lodewijk** en **Naomi** bedanken voor hun inspanning. Ik vond het leuk om jullie te begeleiden en jullie wegwijs te maken in de onderzoekswereld. Heel veel succes met jullie verdere carrière! **Naïla**, succes met je coschappen! Goed dat je je ook door de SPSS- en R-struggles hebt weten te worstelen en een mooie thesis hebt gemaakt.

Ik wil ook de secretaresses bij de interne geneeskunde bedanken voor het organiseren van afspraken en voor het regelen van allerlei logistieke zaken. Dankjulliewel **Mariëlle, Caroline, Saida en Edith**!

Sinds februari 2020 ben ik met veel plezier aan het werk als dokter bij de interne geneeskunde in het Ikazia ziekenhuis. Initieel als ANIOS, maar inmiddels ook al eventjes als AIOS. Ik wil de hele opleidingsgroep internisten, MDL-artsen, longartsen en cardiologen bedanken voor mijn fijne en leerzame tijd in het Ikazia. Beste **dr. de Jongh**, bedankt voor uw vertrouwen dat ik nog voor het lopen van mijn oudste coschap een baan als ANIOS had gekregen in het Ikazia ziekenhuis en dat ik nog geen jaar later als AIOS aan de slag mocht. Beste **drs. Wabbijn**, lieve Marike, wat ben jij een alleskunner! Ik leer verschrikkelijk veel van je en wil je enorm bedanken voor je steun de afgelopen maanden. Ook wil ik **dr. Dees**, mijn mentor, bedanken voor zijn wijze lessen en betrokkenheid. Ik vind het jammer dat u met pensioen gaat. En dan mijn collega arts-assistenten, lieve **Kaziaantjes**, wat heb ik het fijn met jullie! Ik ga (bijna) elke dag met veel plezier naar mijn werk.

Dankjewel **Jelka** voor je adviezen, gezelligheid, discussiemomenten en kopjes thee. De komende vijf jaar samen worden vast net zo leuk als nu!

Op 9 februari zal ik bij worden gestaan door twee hele dierbare vriendinnetjes.

Lieve **Romy**, via de APOP studie leerde ik je kennen en je bleek toen al bijzonder gemotiveerd in het doen van onderzoek. Ik vind het geweldig dat je ook de research master bent gaan doen en ook binnen de acute interne geneeskunde het onderzoek doet. Hoewel ik in de begin periode misschien jou het een en ander heb geleerd, is het de laatste tijd steeds vaker andersom. Onze samenwerking is erg prettig. Ik geniet van onze afspraakjes en uiteraard van je heerlijke kookkunsten. Ik kijk uit naar de verdere samenwerking, naar het moment dat jij op deze plek staat en je onderzoek verdedigt, waar ik achter jou zal staan.

Lieve **Daphne**, vriendinnetje uit het verre oosten (van Nederland). Het is fijn om ook een hele goede vriendin te hebben buiten de Geneeskunde. Je kon me echt af en toe weer met beide benen op de grond zetten en laten zien dat er meer is dan de binnenkant van een ziekenhuis. Fijn ook om (in ieder geval een aantal jaar) een vriendinnetje uit het verre oosten dichtbij in het westen te hebben wonen. De vele sushi- cocktail-, wijn-dates hebben zeker bijgedragen aan dit boekje. Gaan we snel weer een reisje boeken?

Lieve middelbare schoolvriendinnetjes, a.k.a. 7girls4ever, **Fleur**, **Suzan**, **Karien**, **Jolien** en **Luciënne**. Dit is dus de reden dat ik niet veelvuldig in Haaksbergen en omstreken te vinden was. Misschien dat het nu wel gaat lukken om Sinterklaas rond de juiste datum te vieren in plaats van de SinterklaasKerstPaashaasenNieuw meetings van de afgelopen jaren. Ik waardeer het dat jullie de reis naar de mooiste stad van Nederland maken om bij de verdediging aanwezig (en het feestje) te kunnen zijn.

Dan de EMC buddies:

Lieve **Rianne**, vanaf het eerste jaar van de studie Geneeskunde zijn we al acute buddy's. Samen met jou en een aantal andere enthousiastelingen hebben we tijdens onze studie weer leven geblazen in de Spoedeisende Hulp Studenten Organisatie (SEHSO) en veel interessante lezingen en workshops georganiseerd. Ik koester de vele herinneringen aan dat we ons lichaam en geest 'verkochten' aan de wetenschap (lees: bijv. echo-model waren in ruil voor congresdagen). Hoewel we allebei een hele andere hoek binnen de acute geneeskunde interessant vinden, heb ik genoten van je enthousiasme en onze gekke tripjes naar Londen, Chicago en Italië. Hopelijk gaan we in de toekomst vaker dit soort tripjes maken.

Lieve **Nyrée**, inmiddels heb je ook bedacht dat interne geneeskunde eigenlijk heel leuk en interessant is. En nee, het is helemaal niet suf om dat toe te geven! Ook met jou heb ik een leuk tripje gemaakt. Door jouw streep-mee-met-Nyrée-methode hadden we in sneltreinvaart de highlights van Málaga gezien. Laten we dat nog eens herhalen en laten we de komende zomers vaker gaan suppen met onze planken. Dankjewel voor je gezelschap!

Lieve **Karlijn**, dankjewel voor je support de afgelopen maanden. Ik geniet elke keer weer van onze afspraakjes; of het nou met een mok thee op de bank is, een spelletjesavond of een gastronomisch diner is. Op naar nog meer potjes *Pandemic, Ticket to Ride, 30 seconds* en nog veel meer!

Meiden, tien (!!!) jaar geleden begonnen we met studeren en diezelfde tijd zijn we al vriendinnetjes. Time flies when you are having fun. Ik hoop dat we nog heel lang vriendinnen blijven.

Ik prijs mij gelukkig met heel veel lieve **vrienden van de KRPH**. De donderdagavond is vaste prik een avondje ontspanning, waar ik even het onderzoek kon loslaten. Dank jullie wel voor jullie interesse, muzikale afleiding en ook voor jullie support als ik even in de klaagmodus zat. Hopelijk zijn er straks weer veel concerten mogelijk met ons mooie cluppie! En vrienden en vriendinnetjes van de diverse nevenclubs (KRPH cocktailclub, Rotterdampas), laten we streven naar nog meer gezellige momenten!

Ook bedankt aan de vrienden van het **Zwemcentrum Rotterdam**. Wie had dat gedacht? Anniek fanatiek aan het sporten. Dit is onder meer te danken aan de leuke mensen die ik de afgelopen jaren heb ontmoet.

Op de achtergrond zijn er een aantal familieleden geweest die mijn heel erg gesteund hebben. Enkelen daarvan zijn er helaas niet meer. Een daarvan is mijn **opa Brink**, mijn allergrootste fan die tot op hele hoge leeftijd zeer betrokken en geïnteresseerd is geweest in mijn carrière, maar ook in mijn hobby's. Mijn muzikaliteit heb ik ongetwijfeld van u geërfd! Samen met **oma Brink** heeft u gezorgd voor zulke mooie herinneringen.

Lieve **Jos**, je was een hele lieve vrouw voor Pim en moeder voor Maarten, Bouke en Thijs. Ik bewonder je creativiteit; voor elk feestje werd gegarandeerd een nieuwe liedtekst door jou bedacht, maar ook werd er door jou getekend en gehaakt.

Lieve **Pim**, hoe vaak je wel niet gevraagd hebt wanneer ik nou zou gaan promoveren. Meerdere keren heb je herhaald dat je bij mijn promotie zou willen zijn, en hoe oneerlijk dat dat ons beiden niet gegund is. Ik wil je bedanken voor al je steun, al je klushulp, al je taxiritten, al je dinertjes volgens zondags recept inclusief crème brûlée met individuele gasbrandertjes. In gedachten ben je overal waar wij zijn. Ik mis je.

Lieve **Bouke, Erinn, Thijs, Peggy en Maarten,** ik ben trots op ons. Het afgelopen halfjaar was niet leuk. We zijn er echter wel heel sterk en hecht door geworden, en maken er het met elkaar het beste van. Ik ben blij dat ik onderdeel mag zijn van de familie SchefferT.

Opa en oma van Kleef, bedankt voor jullie liefde, gezelligheid en betrokkenheid. Bij een bezoek aan Brielle kan en mag ik nooit met een lege maag vertrekken. Ik geniet van de momenten met jullie. Ik hoop dat jullie gezondheid het toe laat dat jullie bij de promotie plechtigheid aanwezig kunnen zijn.

Lieve **Jorn**, lief groot broertje, dankjewel voor je hulp en interesse. Voor je hulp bij het vinden van een computer, waar ik inmiddels al vele uren achter heb gezeten voor het corrigeren van dit proefschrift. Ook voor je hulp bij het maken van R-codes, voor het meekijken met de stellingen en het nakijken van stukken tekst. Je bent een kanjer en ben super trots dat je bij de plechtigheid aanwezig bent.

Lieve **mam en pap**, a.k.a. de liefste ouders van de hele wereld, ik weet niet hoe ik jullie voldoende kan bedanken. Jullie staan altijd voor me klaar en steunen me onvoorwaardelijk. Jullie hebben me geleerd om door te zetten en hebben mij altijd de mogelijkheid en ruimte gegeven om me te ontwikkelen tot de persoon die ik nu ben. Ik vind het fijn om bij jullie thuis te komen (al is Haaksbergen wel heel ver vanaf Rotterdam/Rhoon). Mam, bedankt voor je luisterend oor, adviezen en knuffels op zijn tijd. Speciale dank aan papa, die mijn enthousiasme deelt en ongelofelijk veel tijd heeft gestoken in zowel de inhoud van dit boekje als de opmaak en cover. Met jou ben ik op mijn zestiende begonnen met het uitvoeren van onderzoek, wat zelfs heeft geleid tot twee posterpresentaties. Bedankt voor de samenwerking met het MST bij het CLEARED-artikel. Je bent een voorbeeld voor me en ik vind het leuk dat ik een beetje in je voetsporen ben getreden.

De afgelopen jaren heb ik heel veel tijd gestoken in het uitvoeren van onderzoek, het schrijven van artikelen en het samenstellen van dit proefschrift. Daarnaast moest ik ook de coschappen en de research master *Clinical research* afronden en begon ik als kersverse dokter. Deze hectiek had ik niet aangekund als er thuis geen stabiele factor was. Lieve **Maarten**, je bent mijn steun en toeverlaat. Vanaf het eerstejaars weekend tijdens de studie Geneeskunde was het al raak, en het is nog steeds fijn met jou! Dankjewel voor je geduld en onvoorwaardelijke liefde. Dat we nog maar veel leuke momenten (o.a. reisjes) met elkaar mee mogen maken. *I love you!*

Anniek

Appendices

About the Author

Anniek Brink was born on June 9th, 1992 in Groningen, the Netherlands. She grew up in a loving family in Haaksbergen, a village near Enschede. She has a broad interest in music, horse riding and scuba diving.

Anniek has always been interested in clinical research. During the final years of her secondary education at the Assink Lyceum in Haaksbergen, she conducted research on the occurrence of catheter-related sepsis in dialysis patients. For her graduation project, she published two posters, one of which was presented at the Dutch "Nefrologiedagen" of 2009, and the other one was presented at the Renal Week of the American Society of Nephrology (2010).

In order to earn admittance to medical school she enrolled at the Winford school in Utrecht and graduated cum laude. In September 2011 she started medical school at the Erasmus University Rotterdam and subsequently participated in research at the department of Internal Medicine under supervision of prof.dr. R. Zietse and dr. J. Alsma.

Preceding her medical school she volunteered at a hospital in Ghana, where she became passionate about working abroad. From 2013 to 2014 she was the president of the STOLA Foundation, a foundation that organizes internships in developing countries for Erasmus MC medical students. She herself did her surgery internship at the Sint Elisabeth Hospital in Curaçao.

Together with a group of enthusiastic students she re-established SEHSO Rotterdam, organising extracurricular courses, lectures and workshops for students interested in emergency medicine.

In 2014, Anniek was accepted by the Netherland Institute of Health Sciences to study Clinical Research. As part of the program she did two in depth epidemiology courses at the Cambridge Institute of Public Health, Cambridge, UK and followed two courses at the Harvard School of Public Health in Boston, Massachusetts, USA during the summer of 2015.

She started her PhD concurrently with the start of the Clinical Research master under supervision of prof. dr. R. Zietse, prof. dr. S.C.E. Klein Nagelvoort-Schuit, prof.dr. H.F. Lingsma, and dr. J. Alsma, which resulted in this thesis.

She graduated as a Master of Science in the fields of Medicine and Clinical Research in 2019.

During the completion of her thesis she started working as a resident internal medicine not in training at Ikazia Hospital, Rotterdam. As of January 2021, she started with her residency in internal medicine at Ikazia Ziekenhuis, Rotterdam, under supervision of drs. Marike Wabbijn and dr. Adrienne Zandbergen (Erasmus Medical Center Rotterdam). After her PhD defence she will continue to be involved in the research of acute medicine. She lives in Rhoon, together with her partner Maarten and their two cats Saar and Pip.