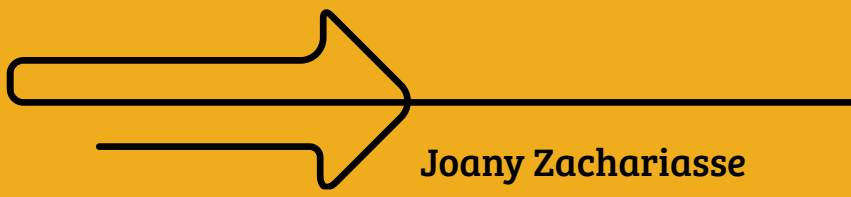
Paediatric Emergency Triage

Improving the initial assesssment of children in the emergency department using electronic health record data



Joany Zachariasse

Paediatric emergency triage

Improving the initial assessment of children in the emergency department using electronic health record data

Joany Zachariasse

COLOFON

ISBN: 978-94-6421-296-9

Cover design: Wendy Bour-van Telgen, Ipskamp Printing Enschede
Layout: Douwe Oppewal, www.oppewal.nl
Printing: Ipskamp Printing Enschede

The study described in Chapter 10 of this thesis was financially supported by Stichting de Drie Lichten and Stichting Sophia Kinderziekenhuis Fonds.
Publication of this thesis was supported by Erasmus University Rotterdam.

PAEDIATRIC EMERGENCY TRIAGE

Improving the initial assessment of children in the emergency department using electronic health record data

Triage van kinderen op de spoedeisende hulp

Het verbeteren van de eerste beoordeling met gegevens uit het elektronisch patiëntendossier

Proefschrift

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

Prof.dr. F.A. van der Duijn Schouten

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 11 mei 2021 om 13.00 uur

door

Johanna Maria Zachariasse
geboren te Vlissingen

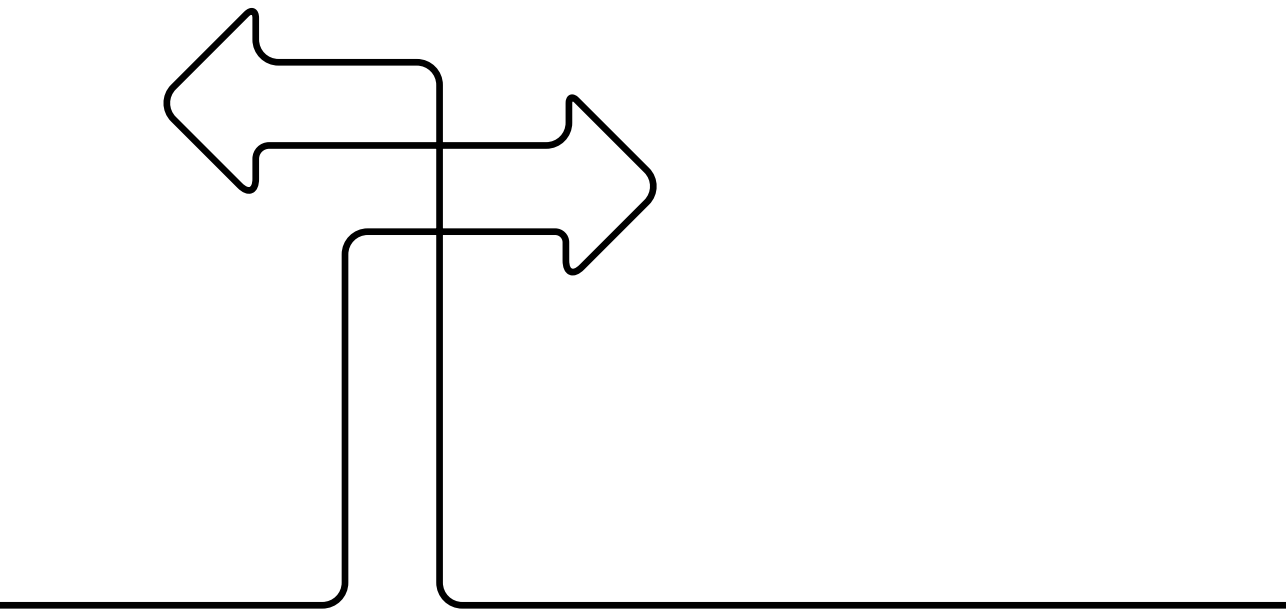
Erasmus University Rotterdam

Promotiecommissie

Promotoren: Prof.dr. H.A. Moll
Prof.dr. J. van der Lei

Overige leden: Prof.dr. M. de Hoog
Prof.dr. E.H.H.M. Rings
Prof.dr. E.W. Steyerberg

TABLE OF CONTENTS



Chapter 1

General introduction

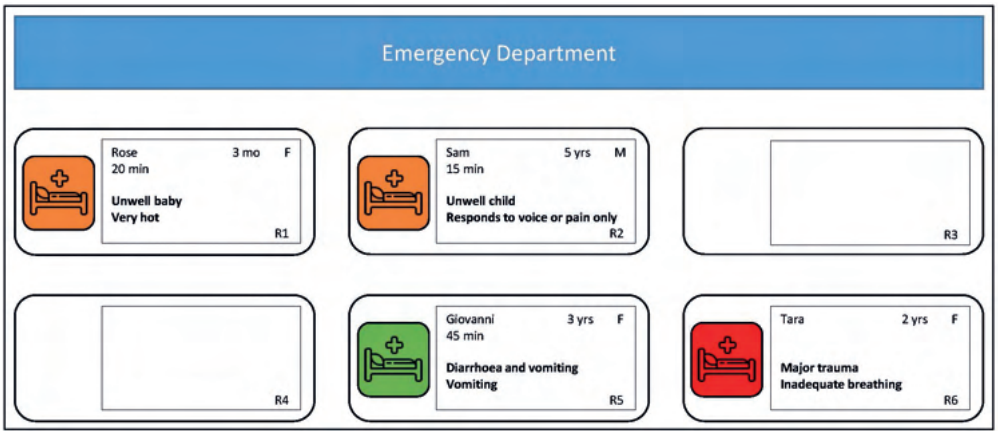
INTRODUCTION

Children account for roughly 20 percent of all attendances to the emergency department (ED).[1, 2] They present with diverse problems, ranging in severity and stage of disease course. Some children present with a novel complaint, while others present with an exacerbation of a known health problem or with multiple complex conditions.

ED utilization is largely unpredictable and the number of children and the complexity of their problems varies from day to day and from hour to hour. Frequently, not all children can be seen immediately by a healthcare professional. Therefore, a system for prioritization needs to be in place to ensure that children as well as adults are seen in order of clinical need instead of order of attendance.[3, 4]

Recognizing the child with serious illness or at risk of deterioration, however, is a major clinical challenge. Children's presenting signs and symptoms are often nonspecific, and characteristic changes in vital signs that signal deterioration generally occur late in the disease course.[5, 6] Moreover, serious illness in children is relatively rare, and the child with serious illness must be identified amidst a much larger population of relatively well children with mild and self-limiting illnesses. To ensure safety at the ED, it is crucial that those severely ill children are identified early and accurately, to avoid harm by delays in treatment.[7, 8] (Figure 1)

Figure 1. Emergency department dashboard illustrating the practice of triage



Example of an ED setting where multiple children have arrived at the same time. The child triaged red requires cardiopulmonary resuscitation and needs immediate attention. Two children were triaged orange and need to be seen by a physician within ten minutes. Although these children are classified as high urgent, this does not necessarily mean that their condition is of high severity. During a quick physician assessment, the child with fever appears well and is has ceased grunting after a dose of paracetamol. Therefore, the somnolent child receives a complete physician consultation first. The infant with vomiting and diarrhea did not exhibit any alarming signs and symptoms and was triaged green. He should be seen within 2 hours by a physician. In some EDs, fast track systems are implemented and patients triaged to the low urgency categories can be seen by general practitioners or dedicated nurses.

Triage systems in the emergency department

Triage is derived from the French word "trier" which means to sort. Triage systems are classification systems, used in emergency departments as a quick assessment to prioritize patients and ensure they are seen in order of clinical priority, rather than in order of attendance.[9] Several triage systems have been developed to standardize the approach to triage, including the Canadian Triage and Acuity Scale (CTAS), the Emergency Severity Index (ESI) and the Manchester Triage System (MTS).[10-13]

The MTS is predominantly used in European countries. It consists of 52 flowcharts, covering patients' chief presenting complaints such as "Headache", "Shortness of breath" and "Wounds". Each flowchart in turn consists of additional signs and symptoms, named discriminators, such as "Airway compromise", "Severe pain" or "New neurological deficit", which are ranked by priority.

In practice, a designated triage nurse selects for each patient the most appropriate flowchart and consequently gathers information on the discriminators from top to bottom. Selection of the first positive discriminator allocates the patient to the consequent urgency level, ranging from immediate (0 minutes maximum waiting time) to non-urgent (240 minutes maximum waiting time) (Figure 2).[14]

Several studies showed good inter-rater reliability between trained nurses for determination of the urgency level.[15, 16]

The performance of triage systems in clinical care

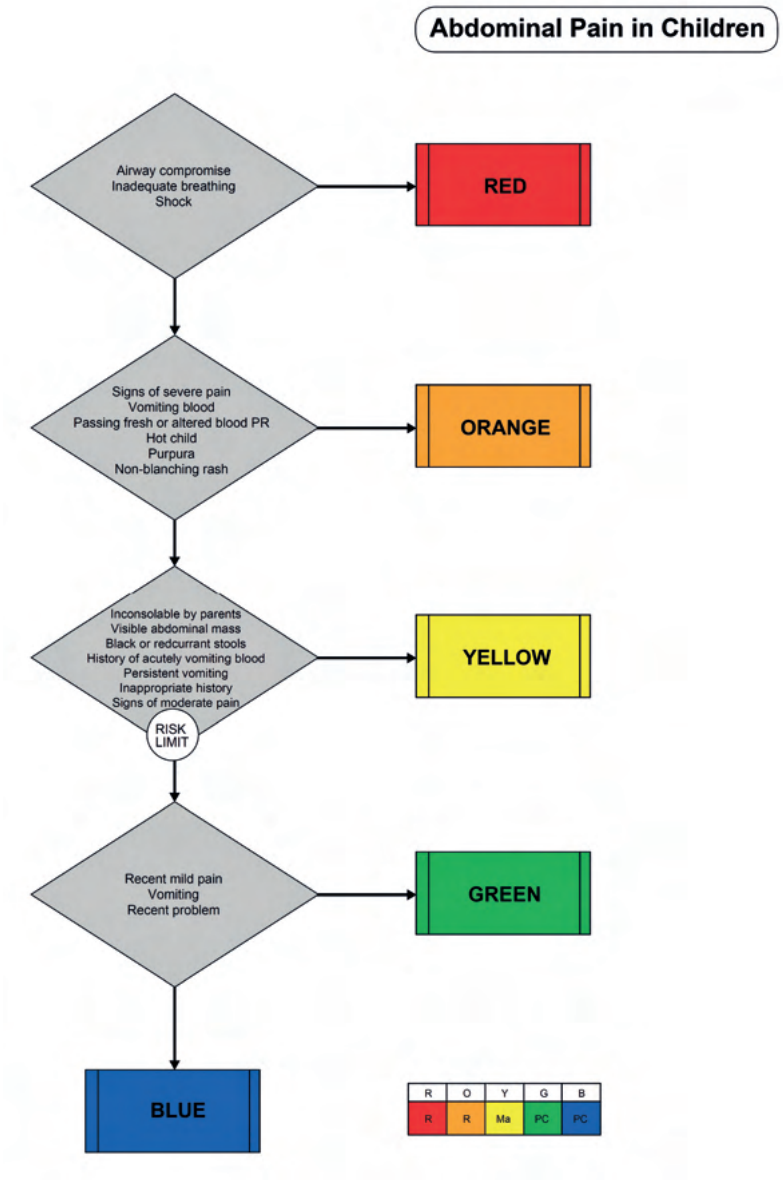
The vast majority of emergency departments in high-income countries use a triage system to prioritize patients. Available triage systems were generally developed based on expert opinion, often as a solution to local challenges in the prioritization of patients. But despite their widespread implementation, research on the performance of these systems in children is limited.[17]

An effective triage system has low undertriage (incorrectly classifying high urgent patients as low urgent), with limited overtriage (incorrectly classifying low urgent patients as high urgent). Undertriage causes delays in treatment of seriously ill patients, potentially leading to morbidity or even mortality. Overtriage makes triage systems less efficient and obstructs the flow for truly urgent patients with consequent delays in treatment and quality of care.

Research on the performance of triage systems is important, both to understand the performance of currently used triage systems and to enable the evaluation of modifications aimed at improvement.

Triage systems are implemented in a wide variety of hospitals with large differences in patient volume and case mix, and within countries with diverse health care systems. Multicentre and international studies are needed to understand the performance of triage systems in these different settings. Moreover, adequate performance measures are required to take into account the multiple ordinal urgency categories of triage systems.

Figure 2. Example of an MTS flowchart



Reprinted with permission from John Wiley & Sons, Ltd (Mackway-Jones K. Manchester Triage Group. Emergency Triage, 3rd edition. London: John Wiley & Sons, Ltd; 2014)

Improving the Manchester Triage System for children

To improve the recognition of critically ill children at the emergency department, existing triage systems can be modified, or new tools can be introduced in the triage assessment.

Because of their widespread implementation, improvements of triage systems have the potential to directly impact clinical care.

Triage systems can be altered by changing the existing structure, for example by moving a certain discriminator to a higher or lower urgency level. Additionally, new predictors of urgency can be added. Examples of potential new predictors include age, gender and vital signs.

The MTS does not require routine measurement of physiological parameters and thus, adding vital signs to the triage system appears a promising way to improve the identification of high urgency patients. In several studies, vital signs appeared early markers of deterioration in hospitalized patients[18-20], and predictors of disease severity in febrile children[21, 22]. Moreover, a study has shown that children with severe undertriage by the MTS often have abnormal vital signs.[23]

New predictors and tools in the initial assessment of children at the emergency department

Other promising predictors of patient urgency, not included in existing triage systems are nurses' clinical impression that a child is ill, and low blood pressure. It is believed that nurses' first impression can play an important role in the identification of patients with serious conditions. There is a vast amount of qualitative literature on judgement and decision-making in nursing practice, but quantitative data on the diagnostic accuracy of nurses' clinical impression in the ED is lacking.[24, 25] Blood pressure plays an important role in the first assessment of adults at the ED, but in children, no consensus exists on the value of routine blood pressure measurements.[26, 27] Moreover, normal blood pressure values vary with age, and accurate age-related reference values for use in the ED are missing.[28]

Besides triage systems, other tools are available to assist in the early recognition of critically ill children. Paediatric Early Warning Scores, also called PEWS, consist of a combination of multiple physiologic parameters. Each of the parameters is assigned a score based on their deviation from the normal, and the individual scores are summed into a final score. PEWS were originally developed for use in hospitalized children to predict deterioration by repeatedly measuring scores and observing trends over time. Several of these scores are now being used in emergency departments to aid triage nurses in the recognition of children that are sick or at risk of deterioration. PEWS are objective measures, do not require spoken language, and do not require any specific training to be applied by healthcare workers. Most currently available PEWS, however, were constructed for use in hospitalized patients and none were developed based on real-world data. Therefore, there is the need for a PEWS that is optimised for use in the ED.

The TRIAGE project: a prospective observational study to improve triage for children in Europe

The TRIAGE project (TRiage Improvements Across General Emergency departments) aims to optimize the triage of children at the ED through a large multicentre prospective observational study based electronic health record data.

Five hospitals from four European countries participate in the project: Erasmus MC-Sophia Children's Hospital, and Maasstad Hospital (the Netherlands), St. Mary's Hospital (United Kingdom), Hospital Fernando da Fonseca (Portugal), and General Hospital Vienna (Austria). The study sites include university and non-university affiliated hospitals of various sizes and in countries with different healthcare systems. Therefore, the study population entails a broad spectrum of children seeking emergency care, generalizable to the majority of emergency departments in Western Europe. All consecutive children under the age of 16 attending the emergency department are included. The project is based on observational data that is routinely collected during emergency department visits. These include demographics, information about triage, signs and symptoms, diagnostics and interventions, final disposition and certain timestamps. The data is automatically extracted from patients' electronic health records and pseudo-anonymized. To ensure quality, site visits were conducted at the start of the study and completeness and accuracy of the data was assessed. Before the analyses, data were checked for quality and outliers, and harmonized.

A fundamental problem in previous studies validating triage systems is the lack of consensus about the outcome measure. Triage systems aim to classify patients based on the urgency of their presenting condition. There is, however, no single outcome measure that captures this concept.[17] Moreover, triage systems typically classify patients into five urgency categories. Dichotomous outcome measures do not capture these different levels and therefore a multilevel reference standard should be used.[29] An important aspect of the TrIAGE project was the development of a reference standard that serves as a proxy of patient urgency.

Objective and outline of the thesis

This thesis aims to improve the first assessment of children presenting at the emergency department.

Therefore, the main objectives are:

1. To evaluate the performance of existing triage systems for the identification of high and low urgency children in the ED
2. To provide recommendations on the performance measures used to compare triage systems
3. To develop new discriminators based on vital signs that improve the Manchester Triage System for children

4. To identify other predictors for urgency in children, including two single predictors (nurses' clinical impression and hypotension)
5. To develop and validate a PEWS based on real-world data, designed for use in the ED

Part I of this thesis addresses the performance of currently used triage systems. **Chapters 2 and 3** describe two studies on the performance of the Manchester Triage System, the most widely used triage system in Europe, in children. A systematic review in **Chapter 4** provides an overview of the available evidence on the performance of triage systems in emergency care for both adults and children. In **Chapter 5,** performance measures for the assessment of modifications of triage systems are evaluated.

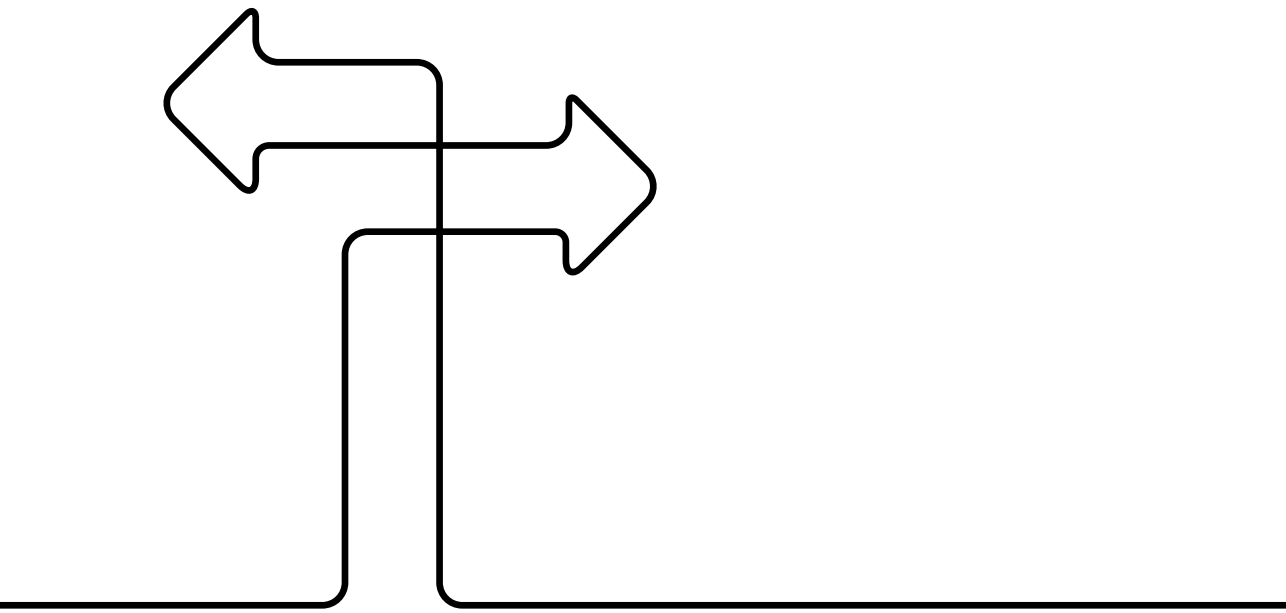
Part II and III of this thesis explore how the initial assessment of children at the emergency department can be improved. Part II focusses on improving the MTS. **Chapter 6** describes a study on the development and performance of vital signs-based modifications to the MTS.

In Part III of this thesis, additional tools in the first assessment of children are addressed. In the first chapters, single predictors of urgency in children are studied. **Chapter 7** determines the diagnostic accuracy of nurses' clinical impression that a child appears ill, while **Chapters 8 and 9** describe the normal ranges for low blood pressure and its value in the recognition of serious illness in children. In **Chapters 10** a novel PEWS for use in the ED is developed and validated. This ED-PEWS is validated for children with underlying chronic conditions in **Chapter 11**.

To conclude, the discussion in **Chapter 12** of this thesis elaborates on the main findings in this thesis and addresses future perspectives. In **Chapter 13** a summary is provided.

PART I

The performance of triage systems



Chapter 2

Safety of the Manchester Triage System to detect critically ill children at the emergency department

Joany M Zachariasse, Jan Willem Kuiper, Matthijs de Hoog, Henriëtte A Moll, Mirjam van Veen

J Pediatr. 2016 Oct;177:232-237

ABSTRACT

Objective To assess the safety of the Manchester Triage System in pediatric emergency care for children who require admission to the intensive care unit (ICU).

Study design Between 2006 and 2013, 50,062 consecutive emergency department visits of children younger than the age of 16 years were included. We determined the percentage of undertriage, defined as the proportion of children admitted to ICU triaged as low urgent according to the Manchester Triage System, and diagnostic performance measures, including sensitivity, specificity, and diagnostic OR. Characteristics of undertriaged patients were compared with correctly triaged patients. In a logistic regression model, risk factors for undertriage were determined.

Results In total, 238 (28.7%) of the 830 children admitted to ICU during the study period were undertriaged. Sensitivity of high Manchester Triage System urgency levels to detect ICU admission was 71% (95% CI 68%-74%) and specificity 85% (95% CI 85%-85%). Severity of illness was lower in undertriaged children than correctly triaged children admitted to ICU. Risk factors for undertriage were age <3 months, medical presenting problem, comorbidity, referral by a medical specialist or emergency medical services, and presentation during the evening or night shift.

Conclusion The Manchester Triage System misclassifies a substantial number of children who require ICU admission. Modifications targeted at young children and children with a comorbid condition could possibly improve safety of the Manchester Triage System in pediatric emergency care.

INTRODUCTION

Triage systems are used in emergency departments (EDs) to prioritize patients and to ensure that they are seen in order of clinical need when demand exceeds capacity. In Europe, the Manchester Triage System (MTS) is the most frequently used emergency medical triage system.[14] The MTS is a flowchart-based algorithm, that classifies patients into 1 out of 5 urgency categories, each corresponding to a predetermined maximum waiting time.

Although the MTS is used widely, research evaluating its safety for the triage of children is limited. The safety of a triage system refers to its ability to identify high-urgent patients. Misclassification of high-urgent patients to a low-urgency level, so-called "undertriage", causes delay in the care of severely ill patients and potentially leads to morbidity or even mortality. Children, accounting for more than 25% of the workload of EDs, are at increased risk of undertriage: they suffer from a different spectrum of disease than adults, they frequently present with nonspecific complaints, and characteristic changes in vital signs that signal deterioration in adults often occur late in the disease course.[5, 6] Two previous studies assessed the diagnostic accuracy of the MTS in children and concluded that validity of the MTS for the triage of children was moderate[30, 31]; however, these studies did not specifically address safety of the MTS for high-urgent children, nor did these studies determine predictors of undertriage.

Admission to the intensive care unit (ICU) is a specific and clinically relevant outcome to study the safety of triage systems.[32] Patients admitted to the ICU are by definition either critically ill or at risk of developing life-threatening conditions. Moreover, delays in admission to the ICU have been shown to negatively impact health outcomes in adults.[33] We propose as minimum requirement for a triage system that it accurately identifies patients in need of admission to the ICU. Therefore, we performed a large observational study to determine the safety of the MTS in pediatric emergency care for children who require admission to the ICU. Moreover, we aimed to describe the group of undertriaged children and identify risk factors for undertriage.

METHODS

We evaluated the safety of the MTS as part of an ongoing study on the validity of the MTS in children.[23, 31, 34, 35] The Medical Ethics Committee of the Erasmus MC approved the study, and the requirement for informed consent was waived.

Erasmus MC-Sophia Children's Hospital is an urban university hospital in the city of Rotterdam, the Netherlands. The pediatric ED serves the inner-city population but also holds a regional function for patients with significant comorbidity. Approximately 7000 children are seen yearly. Major trauma cases are diverted to the adult Erasmus MC ED.

The pediatric ICU is a tertiary medical and surgical unit with approximately 1500 planned and unplanned admissions yearly. In addition to the patients who are admitted from the Erasmus ED, the ICU receives a large proportion of its patients from regional hospitals.

We included all consecutive ED visits of children younger than the age of 16 years at the Erasmus MC-Sophia Children's Hospital between January 1st, 2006, and December 31th,2012. We excluded patients with a tracheal cannula or home care ventilation because these patients cannot be admitted to the general wards of the hospital for logistic reasons and therefore may have other reasons for admission to the ICU than severity of illness.

Admission to the ICU was defined as admission to the ICU immediately after a visit to the ED. Children who were admitted to the ICU after first being admitted to the general ward, for example, due to clinical deterioration, were not classified as ICU admissions in the study. Indications for admission to the ICU conform to national standards and include acute or threatening failure of 2 or more organ systems; requirement of advanced respiratory support, expected to last >24 hours or in a child younger than 1 year of age; or need for intensive monitoring because of acute or threatening failure of 1 or more organ systems.[36] Comorbidity alone is no indication for admission to our ICU.

Triage at the Erasmus MC-Sophia Children's Hospital was performed by ED nurses trained in the MTS. A computerized version of the official Dutch translation of the MTS was used, with validated modifications for febrile children implemented from April 2007 onwards.[35, 37] Nurses routinely recorded data of all ED visits on structured electronic forms, during or shortly after the ED visit. These forms contain items regarding patient characteristics, vital signs, working diagnosis and follow-up.

Data on admission to the ICU, including length of stay, mortality, and severity-of-illness scores, were retrieved from electronic medical ICU records. These data were collected routinely as part of the pediatric intensive care evaluation, a national pediatric ICU registry for benchmarking and research purposes.[38] We quantified severity of illness with the Pediatric Risk of Mortality Score (PRISM) 3, for which the greater scores indicate greater risk of mortality (maximum score 74) and the Pediatric Index of Mortality (PIM) 2, for which the score (percentage) indicates the predicted death rate.[39, 40]

To assess comorbidity, one investigator reviewed all undertriaged (low-urgent, ICU-admitted) patients and a random sample of correctly triaged low-urgent non-ICU admitted patients and recorded all underlying chronic conditions based on the written information available in the patients' medical records, blinded to information on MTS urgency classification. Chronic diseases were classified according to the Pediatric Medical Complexity Algorithm into complex chronic disease, noncomplex chronic disease and no chronic disease.[41] Children are defined as having a complex chronic condition if 2 or more body systems are affected, if they suffer from a progressive condition or a malignancy, or if they are continuously dependent on technological support.

Data analysis

Because we had little missing information on triage classification or outcome (5%), we used a complete case analysis. Demographic and clinical characteristics of included patients were presented as proportions or medians and IQRs.

We dichotomized MTS urgency categories into high urgent (MTS urgency 1 and 2) and low urgent (MTS urgency 3, 4, and 5). The MTS defines a maximum waiting time before first contact with a physician: 0 and 10 minutes waiting time for urgency levels 1 and 2 and 60, 120 and 240 minutes waiting time for the urgency levels 3, 4 and 5. We set our cut-off between urgency level 2 and 3, because we consider 10 minutes before first contact with a physician a safe time window for patients who require admission to the ICU. MTS urgency 3 has a maximum waiting time of 60 minutes, which can lead to delays in care for critically ill patients. Safety of the MTS was assessed by the percentage of undertriage, defined as the proportion of patients admitted to the ICU who were triaged initially as low urgent. Moreover, we calculated the sensitivity, specificity, predictive values, likelihood ratios and the diagnostic OR of MTS high-urgency classification for the detection of admission to the ICU.

To evaluate whether undertriaged patients were clinically different from correctly triaged patients admitted to the ICU, we compared several measures of severity of illness between these 2 groups: PIM2 and PRISM3 score, length of stay, need of ventilatory support and mortality. Groups were compared by use of the Pearson's χ^2 test for categorical or the Mann-Whitney *U* test for continuous variables.

To identify risk factors for undertriage, multivariable logistic regression analysis was performed to compare the undertriaged patients with the low-urgent patients who were not admitted to the ICU. Predictor variables were selected on the basis of previous research[9,16] and clinical knowledge. We included all candidate predictor variables in the model, independent of their statistical contribution. Age was converted into an ordinal variable with clinically relevant categories (0-<3 months; 3-<12 months; 1-<4 years; 4-<8 years; 8-<16 years). Comorbidity was only available in a sample of patients and therefore the OR was calculated independently.

SPSS version 20.0 (SPSS Inc, Chicago, Illinois) and the VassarStats website (www.vassarstats.net) were used for the statistical analysis.

RESULTS

During the study period, there were 53,180 ED visits of children younger than the age of 16 years. A total of 425 (0.8%) visits were excluded because the patient had a tracheal cannula or home care ventilation, and 2,693 (5.1%) visits had incomplete data (Figure 1). Therefore, 50,062 visits were included in the analysis resulting in 830 (1.7%) admissions to the ICU. The percentage of admissions to the ICU was not statistically different between

the study population and the group of children with incomplete data (χ^2 [1] = 0.24, p=.62). Demographic and clinical characteristics of ED visits are presented in table 1.

Table 1. Demographics of the study population

Variables	Not admitted to the ICU (n = 49,232)	Admitted to the ICU (n = 830)
Age, median (IQR), y	4.1 (1.4 – 9.4)	2.0 (0.3 – 8.0)
Male sex, n (%)	28,758 (58.4)	513 (61.8)
Presenting problem, n (%)		
Medical	34,443 (70.0)	788 (94.9)
Surgical*	14,789 (30.0)	42 (5.1)
Presenting problem, n (%)		
Respiratory or ear, nose, and throat	5338 (10.8)	186 (22.4)
Gastrointestinal	7012 (14.2)	87 (10.5)
Neurologic or general malaise	10,782 (21.9)	329 (39.6)
Other medical	9,893 (20.1)	179 (21.6)
Minor trauma and wounds	16,207 (32.9)	49 (5.9)
Mode of referral, n (%)		
Self	21,303 (43.3)	103 (12.4)
GP	8,481 (17.2)	85 (10.2)
Emergency medical service	3,205 (6.5)	301 (36.2)
Medical specialist	8,919 (18.1)	218 (26.3)
Other / Unknown	7,324 (14.9)	123 (14.8)

GP, general practitioner
*excluding surgical abdominal problems

According to our definition, 238 (28.7%) of the children admitted to the ICU were undertriaged: 176 (21.2%) to MTS urgency 3 and 62 (7.5%) to MTS urgency 4 or 5 (Table 2).

Sensitivity of a high MTS urgency level to detect admission to the ICU in children was 71% (95% CI 68% - 74%) and specificity 85% (95% CI 85% - 85%). The diagnostic OR was 14.1 (95% CI 12.1 - 16.4). Modifications of the MTS that were implemented during the study period had a negligible impact on its sensitivity (Table 3).

Figure 1. Flow diagram of the study population

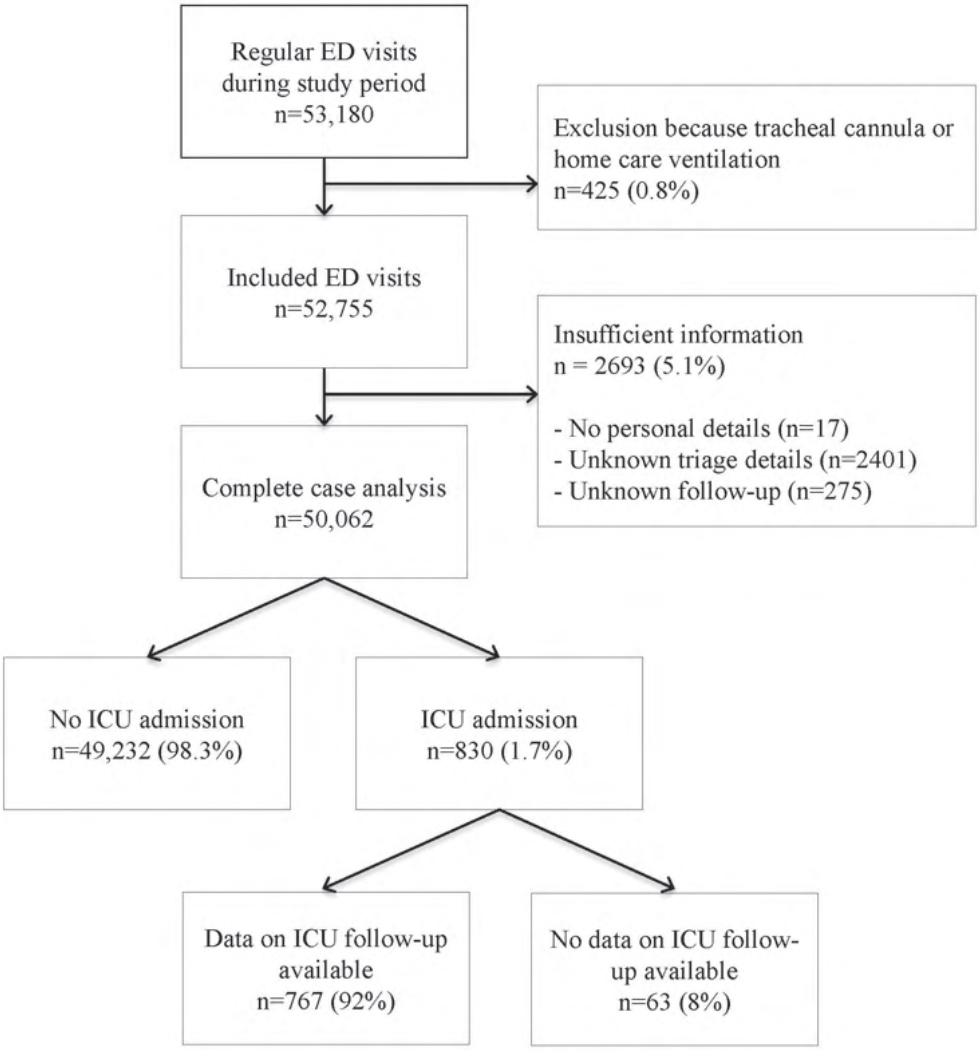


Table 2. MTS urgency categories assigned to the children at triage

	MTS triage category	Not admitted to the ICU (n = 49,232)	Admitted to the ICU (n=830)
High urgent, n (%)	Immediate	749 (1.5)	331 (39.9)
	Very urgent	6,630 (13.5)	261 (31.4)
Low urgent, n (%)	Urgent	22,722 (46.2)	176 (21.2)
	Standard	18,035 (36.6)	59 (7.1)
	Non urgent	1,096 (2.2)	3 (0.4)

Table 3. Diagnostic performance of high urgent MTS categories to identify children requiring ICU admission

Performance measure	Total (n = 50,062)	MTS original[a] (n = 9,020)	MTS with modifications[b] (n = 41,042)
Sensitivity	0.71 (0.68 - 0.74)	0.70 (0.62 - 0.77)	0.72 (0.68 - 0.75)
Specificity	0.85 (0.85 - 0.85)	0.80 (0.80 - 0.81)	0.86 (0.86 - 0.86)
Positive predictive value	0.07 (0.07 - 0.08)	0.05 (0.04 - 0.06)	0.08 (0.07 - 0.09)
Negative predictive value	0.99 (0.99 - 1.00)	0.99 (0.99 - 1.00)	0.99 (0.99 - 1.00)
Positive likelihood ratio	4.76 (4.54 - 4.99)	3.56 (3.17 - 4.00)	5.13 (4.87 - 5.41)
Negative likelihood ratio	0.34 (0.30 - 0.38)	0.37 (0.29 - 0.48)	0.33 (0.29 - 0.37)
Diagnostic OR	14.11 (12.11 - 16.43)	9.53 (6.62 - 13.74)	15.53 (13.13 - 18.37)

[a] MTS original: Jan 2006-May 2007.
[b] MTS with modifications: June 2007-December 2009; In August 2009 the second edition of the MTS was implemented which only contained minor changes compared to the first edition.

Undertriaged patients had significantly lower severity of illness, as measured by PIM2 and PRISM3 severity-of-illness scores, than patients admitted to the ICU that were correctly triaged (0.93 vs 1.26 and 0 versus 2 respectively, both $P<.001$). They also required less-invasive or noninvasive ventilatory support (11.4% versus 35.7%, $P<.001$). In the group of patients with follow-up information available, none of the 210 undertriaged patients died compared with 41 of the 557 (7.4%) correctly triaged patients admitted to the ICU (Table 4). Table 5, a narrative description of the undertriaged patients with a >5% predicted mortality rate according to the PIM score, illustrates the indications for ICU admission in the group of undertriaged children and the complexity of this patient group.

Table 4. Estimates of illness severity for correctly triaged (high MTS urgency) and undertriaged (low MTS urgency) children

Variable	ICU admitted, triaged as high urgent (n = 557)	ICU admitted, triaged as low urgent (n = 210)	*P* value
PIM2, median (IQR), % mortality risk	1.26 (0.87 - 4.41)	0.93 (0.75 - 1.36)	*P*<0.001
PIM2, n (%), mortality risk categorized			
<1%	202 (36.3)	112 (53.3)	*P*<0.001
1 – 5%	228 (40.9)	89 (42.4)	
5 – 15%	62 (11.1)	9 (4.3)	
15-30%	23 (4.1)	0	
≥30%	42 (7.5)	0	
PRISM3, median (IQR), score	2 (0 - 7)	0 (0 - 2)	*P* <0.001
ICU length of stay, median (IQR), d	3 (2 - 5)	2.5 (2 - 4)	*P* =0.141
Ventilation required, n (%)	199 (35.7)	24 (11.4)	*P* <0.001
Mortality in ICU, n (%)	41 (7.4)	0	*P* <0.001

*Slater et al.[16] **Pollack et al.[15]

Table 5. Narrative description of the 9 undertriaged patients with >5% predicted death rate according to the PIM 2 score

Child	Age	MTS urgency	Presenting complaint at triage	Comorbidity	Working diagnosis at end ED visit
1	1 y	3	Fever, a cold	Dilated cardiomyopathy	Impending cardiac decompensation
2	3 mo	3	Dehydration?	Cystic fibrosis, intestinal atresia with reversed jejunostoma	Bilious vomiting and dehydration, due to ileus or distal intestinal obstruction syndrome
3	12 y	3	Episode of unconsciousness and body jerks	T-ALL	Seizures due to hyponatremia
4	14 y	3	Shortness of breath, fever, increased seizure frequency	Intellectual disability, epilepsy, spastic tetraplegia	Status epilepticus and respiratory insufficiency
5	2 y	3	Fall from stairs	-	Head injury, complicated by altered consciousness and localized seizures
6	2 mo	4	Cardiac problems	-	Dilated left ventricle with decreased function
7	3 y	3	Cough, sputum, increased seizure frequency	Intellectual disability, West syndrome	Increased seizure frequency due to upper respiratory tract infection. Need for continuous infusion of midazolam.
8	11 mo	3	Increased seizure frequency	Treatment-resistant epilepsy	Status epilepticus and respiratory insufficiency
9	25 d	3	Strangulated inguinal hernia	-	Strangulated inguinal hernia

T-ALL, T-cell Acute Lymphoblastic Leukemia

The risk of undertriage was greatest in children younger than the age of 3 months (OR 2.87; 95% CI 2.00 - 4.10) and lowest in children aged 1-<4 years (OR 0.61; 95% CI 0.41 - 0.91) and 4-<8 years (OR 0.57; 95% CI 0.36 - 0.90) compared with the oldest age group of 8-<16 years. All medical-presenting problems showed an increased risk of undertriage, compared with surgical-presenting complaints. Referral by a medical specialist or emergency medical services and presentation during evening or night shift also increased the risk of undertriage (Table 6).

The review of medical records of all 238 low-urgent patients admitted to the ICU showed that 137 (58%) had an underlying chronic disease, including 81 (34%) with a complex chronic condition. In contrast, 33.0% of low urgent patients not admitted to the ICU had an underlying chronic condition, which was complex in 20.6% of the cases. Therefore, patients with a chronic condition had a greater risk of being undertriaged (OR 2.8; 95% CI 2.0 - 3.8).

Table 6. Risk factors for undertriage, as determined by univariable and multivariable logistic regression

Determinants	Low-urgent patients (n = 42,091)	ICU admissions (n = 238)	OR, univariable (95% CI)	OR, multivariable (95% CI)
Age				
<3 months	3,064	69	4.89 (3.46 - 6.91)	2.87 (2.00 - 4.10)
3 -<12 months	5,055	40	1.69 (1.14 - 2.52)	1.16 (0.77 - 1.74)
1-<4 years	11,992	41	0.73 (0.49 - 1.08)	0.61 (0.41 - 0.91)
4-<8 years	8,752	26	0.63 (0.40 - 1.00)	0.57 (0.36 - 0.90)
8-<16 years	13,228	62	*Reference*	*Reference*
Gender				
Female	17,636	91	0.86 (0.66 - 1.12)	0.87 (0.67 - 1.14)
Male	24,455	147	*Reference*	*Reference*
Presenting problem				
Respiratory and ENT	3,077	23	6.82 (3.64 - 12.77)	4.69 (2.46 - 8.94)
Gastrointestinal	6,323	51	7.36 (4.25 - 12.75)	4.27 (2.40 - 7.59)
Neurologic and general malaise	8,870	79	8.13 (4.81 - 13.74)	3.95 (2.29 - 6.80)
Other medical	8,420	68	7.37 (4.33 - 12.54)	4.38 (2.52 - 7.59)
Minor trauma and wounds	15,401	17	*Reference*	*Reference*
Referral				
Self	19,259	32	*Reference*	*Reference*
GP	7,004	27	2.33 (1.39 - 3.88)	1.58 (0.93 - 2.66)
Emergency service	1,745	30	10.51 (6.37 - 17.34)	8.45 (5.06 - 14.09)
Medical specialist	7,869	98	7.58 (5.08 - 11.30)	5.55 (3.67 - 8.40)
Other	6,214	51	4.97 (3.19 - 7.74)	3.21 (2.04 - 5.05)
Shift				
Day	20,435	95	*Reference*	*Reference*
Evening	18,962	120	1.36 (1.04 - 1.79)	1.76 (1.33 - 2.32)
Night	2,694	23	1.84 (1.17 - 2.91)	1.78 (1.11 - 2.85)

DISCUSSION

This large observational study demonstrates that a substantial number of critically ill children were not classified as high urgent when triaged by the MTS. Children younger than the age of 3 months, children presenting with medical problems, and children with underlying chronic conditions were at risk of undertriage, as well as children referred by emergency medical services or medical specialists and children presenting during evening and night shifts.

The validity of the MTS has been studied previously in adults and children. Because there is no gold standard to evaluate the validity of triage systems, the majority of published research reports the presence and strength of associations between triage category and a certain outcome measure such as hospitalization or resource use.[32, 42-44] A strong association between triage category and outcome, however, does not guarantee safety if a small but seriously ill group of patients is incorrectly triaged. Only one small study assessed the validity of the MTS for the detection of ICU admission in adults and found a sensitivity of 63%.[32] The study concluded that the MTS is a sensitive tool for the detection of critically ill patients and that most errors are caused when nurses do not apply the system correctly. We argue that this conclusion should be interpreted with caution because true patient urgency was determined by a retrospective assessment of all patient records. Previously, we performed 2 studies in which we assessed diagnostic accuracy of the MTS compared with a 5-level reference standard.[30, 31] The reference standard consisted of a combination of vital signs, presence of potentially life-threatening conditions, ED resource use, and follow-up. These studies found that the MTS has a sensitivity of 63% for the detection of high-urgent patients, and a proportion of undertriage ranging from 12% to 15%. Only one study assessed determinants of severe undertriage and found that young age and use of a general flowchart were risk factors.[23]

This is the first study that specifically assesses the safety of the MTS for the identification of children in need of admission to the ICU. With a large dataset of more than 50,000 patients, we were able to assess safety by the relatively rare event of ICU admission and determine risk factors for undertriage. Because we had few missing data, the risk of selection bias is low.

A limitation of the study is the use of single-center data from a university children's hospital. The results will therefore be primarily generalizable to comparable tertiary care centers. Moreover, we did not include major trauma cases.

We used ICU admission as a proxy for patient urgency to study the safety of the MTS. Patients admitted to the ICU are by definition either critically ill or at risk of developing life-threatening conditions and it is generally not considered safe for these patients to wait for 1 hour, before they are seen by a physician. Furthermore, admission to the ICU is an objective measure that is relatively uniformly applied in different countries and is undoubtedly related to patient severity.

Incorrect triage can be caused by insufficiencies of the triage system itself or by failure of the nurse to apply it correctly. Because all triage nurses in the study received standardized MTS training, and the MTS has very explicit discriminators which always lead to the same urgency level regardless of the flowchart used, it is unlikely that the latter plays a major role. Previous studies evaluating the MTS also have shown good interrater agreement.[15, 45] Performance of triage, however, is likely to be influenced by many factors, including some variability in the application of the triage system, which reflects its performance in clinical practice.

When interpreting these results, we need to consider that the prevalence of admission to the ICU in children attending the ED was low, and therefore the absolute number of undertriaged patients is relatively small compared with the total number of ED visits. Moreover, the subset of undertriaged children was on average less severely ill compared with the subset of correctly triaged patients admitted to the ICU. Also, by design, our study could not determine whether the potential delays in care due to the undertriage resulted in adverse health outcomes. Regardless of this, we argue that our study indicates a weakness of the MTS that needs to be addressed. Five percent of the undertriaged patients, admitted to ICU had a mortality risk of more than 5%, and 12% were in need of ventilatory support, and their median length of stay at the ICU was 3 days. Even though there is clearly a range in severity of conditions that require admission to the ICU, we believe these results indicate that also the subset of less-severe patients admitted to ICU were in need of intensive care. In addition, our study shows that that the group of undertriaged patients mainly consisted of young children and children with comorbidity. These are particular vulnerable subgroups of patients, with a high prevalence of nonspecific signs and symptoms, and an increased risk of unexpected deterioration.[34, 46, 47] Moreover, the notion that outcomes of certain conditions can be improved by early provision of therapy is widely adopted for adults and children,[48, 49] and it has been shown in adults that delayed transfer of critically ill patients from the ED to the ICU is associated with worse patient outcomes.[33] Although there is no consensus when a triage system can be considered safe, we propose as minimum requirement for any triage system that it accurately identifies patients in need of admission to the ICU.

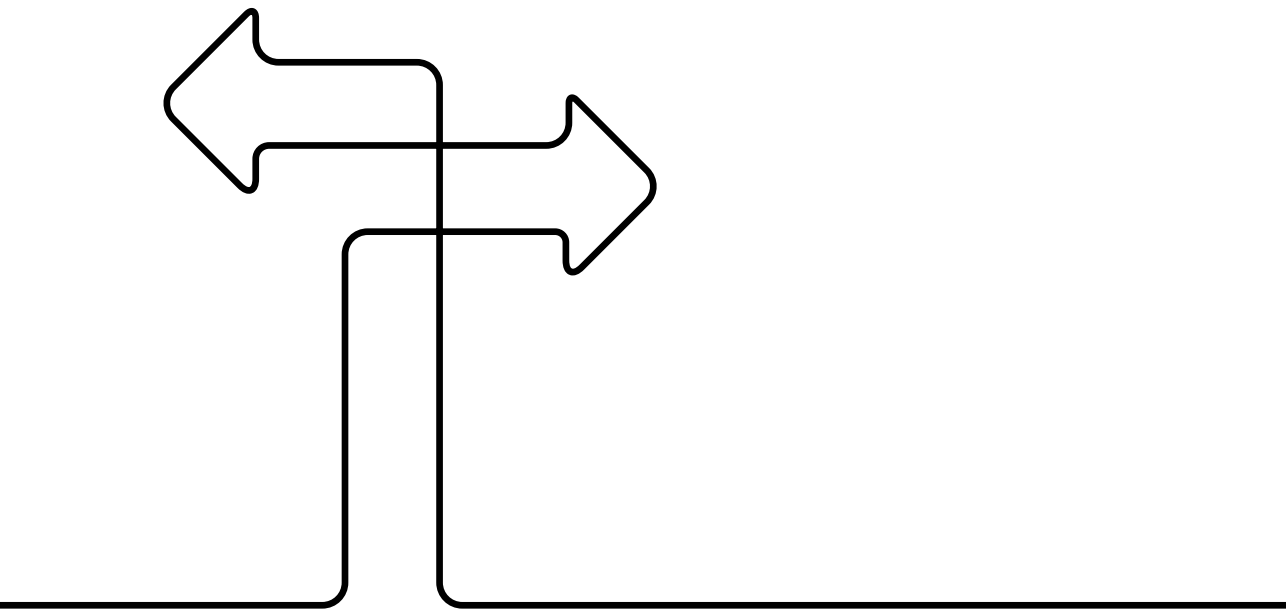
Several factors were found to be associated with undertriage. Patient-related risk factors include age younger than 3 months, medical presenting problem, and underlying chronic condition. Triage of these patient groups is challenging because nonspecific signs and symptoms commonly are present. Referral by emergency medical services or a medical specialist and presentation during evening or night shift were risk factors for undertriage. We hypothesize that medical complexity, differences in patient populations during different times of the day, and ED crowding may have led to underestimation of urgency in these patient groups, which needs further evaluation.

We believe our results are important for clinical practice. First, it is essential to be aware that undertriage of critically ill children at the ED occurs and that certain subgroups

of children, including young children and children with comorbidity are at risk. Second, modifications of the MTS involving patient-related risk factors, may decrease the undertriage of critically ill patients. It is important to note that modifying the MTS, for example, by adding a new discriminator will lead to a reduction in undertriage, at the cost of an increase in overtriage. This means that only predictors that distinguish well between high- and low-urgent children should be added to the system. The determinants found in our study could be a good starting point. Including comorbidity in general as discriminator in the MTS would likely lead to overtriage, but adding discriminators concerning specific types of comorbidity to some flowcharts is a promising modification.[34] The latest edition of the MTS already contains 2 new flowcharts specifically aimed at neonates and young children.[14] Because this latest edition was not available during our study, future research should determine whether these modifications improve triage in the youngest patient groups.

ACKNOWLEDGEMENTS

We thank Idse Visser (Dutch Pediatric Intensive Care Evaluation) for his help in collecting and interpreting the ICU data.



Chapter 3

Validity of the Manchester Triage System in emergency care: a prospective observational study

Joany M Zachariasse, Nienke Seiger, Pleunie P Rood, Claudio F Alves, Paulo Freitas, Frank J Smit, Gert R Roukema, Henriëtte A Moll

PloS One. 2017 Feb 2;12(2):e0170811

ABSTRACT

Objective To determine the validity of the Manchester Triage System (MTS) in emergency care for the general population of patients attending the emergency department, for children and elderly, and for commonly used MTS flowcharts and discriminators across three different emergency care settings.

Methods This was a prospective observational study in three European emergency departments. All consecutive patients attending the emergency department during a 1-year study period (2010-2012) were included. Validity of the MTS was assessed by comparing MTS urgency as determined by triage nurses with patient urgency according to a predefined 3-category reference standard as proxy for true patient urgency.

Results 288,663 patients were included in the analysis. Sensitivity of the MTS in the three hospitals ranged from 0.47 (95%CI 0.44-0.49) to 0.87 (95%CI 0.85-0.90), and specificity from 0.84 (95%CI 0.84-0.84) to 0.94 (95%CI 0.94-0.94) for the triage of adult patients. In children, sensitivity ranged from 0.65 (95%CI 0.61-0.70) to 0.83 (95%CI 0.79-0.87), and specificity from 0.83 (95%CI 0.82-0.83) to 0.89 (95%CI 0.88-0.90). The diagnostic odds ratio ranged from 13.5 (95%CI 12.1-15.0) to 35.3 (95%CI 28.4-43.9) in adults and from 9.8 (95%CI 6.7-14.5) to 23.8 (95%CI 17.7-32.0) in children, and was lowest in the youngest patients in 2 out of 3 settings and in the oldest patients in all settings. Performance varied considerably between the different emergency departments.

Conclusions Validity of the MTS in emergency care is moderate to good, with lowest performance in the young and elderly patients. Future studies on the validity of triage systems should be restricted to large, multicenter studies to define modifications and improve generalizability of the findings.

INTRODUCTION

Triage at the emergency department (ED) aims to prioritize patients when clinical demand exceeds capacity.[9] As the burden on emergency departments worldwide is steadily increasing, triage remains a fundamental intervention to manage patient flow safely and to ensure that patients who need immediate medical attention are timely treated, particularly in case of overcrowding.[50-53]

The Manchester Triage System (MTS) is one of the most commonly used triage systems in Europe.[14] It enables nurses to assign a clinical priority to patients, based on presenting signs and symptoms, without making any assumption about the underlying diagnosis. The MTS allocates patients to one out of five urgency categories, which determine the maximum time to first contact with a physician. Despite its widespread implementation, validity of the MTS remains uncertain. Previous research consists of single center studies,[42, 43, 54, 55] studies restricted to certain age groups or specific medical conditions,[31, 32, 34, 56, 57] and studies analyzing validity by trends in resource use or hospitalisation.[42-44, 54] To date, no study has evaluated performance of the MTS in a large, heterogeneous cohort of patients, at different emergency departments, and with a reference standard that is independent of triage, correlated to severity of illness, and applicable to patients with a wide range of presenting problems.

The aim of this study is to determine the performance of the Manchester Triage System for the general population of patients attending the emergency department and specifically for children and elderly, the most vulnerable groups of patients. Moreover, we aim to evaluate the performance of the most commonly used MTS flowcharts and discriminators. Knowledge about the validity of MTS can provide insight in its performance, it enables the comparison with other triage systems and it can support targeted modifications for improvement.

METHODS

Study design

The study is based on a multicenter prospective observational cohort of patients presenting to emergency departments in three different practice settings. Data collected during routine care was automatically extracted from electronic medical health records. The validity of the MTS was assessed by comparing MTS urgency as determined by triage nurses with patient urgency according to a predefined 3-category reference standard. Moreover, we assessed validity by the ability of the MTS high urgency categories to identify patients requiring Intensive Care Unit (ICU) admission or patients that died at the ED. We evaluated the performance of the MTS for different age groups and for the most commonly used flowcharts and discriminators. The study was approved by

the institutional review boards of all participating institutions and the need for written informed consent from the participants was waived.

Study population and setting

All consecutive patients attending the emergency departments of the Erasmus MC, Rotterdam, the Netherlands (July 2010 to July 2011); Maasstad Hospital, Rotterdam, the Netherlands (July 2011 to July 2012); and Hospital Professor Doutor Fernando da Fonseca (hereafter: Hospital Fernando Fonseca), Lisbon, Portugal (September 2011 to September 2012) were included in the study. Before the study period, all hospitals had two to five years of experience with the MTS.

Erasmus MC is an inner-city university hospital and tertiary care referral and trauma centre, with an ED receiving approximately 24,000 adults and 7,000 children a year. The ED delivers general emergency medicine, but as a tertiary care facility is specialized in complex care. Because the Netherlands has a strong system of primary care, and general practitioners act as gatekeepers, the proportion of low urgent patients is relatively small.

Maasstad Hospital is an inner-city teaching hospital with a mixed emergency department for adult and pediatric patients receiving approximately 38,000 patients a year. The ED delivers general emergency and trauma care. Similarly to the Erasmus, the proportion of low urgent patients is relatively small, because patients with minor complaints are usually seen by the GP or GP cooperative.

Hospital Fernando Fonseca is an inner-city community hospital with an annual census of approximately 190,000 adults and 60,000 children. The hospital delivers general emergency care and trauma care except neuro-surgery. Primary care is often not accessible for patients, and the ED is frequented by a large proportion of patients with minor complaints.

Therefore, settings with a different case-mix contributed to the study.

Manchester Triage System

The MTS is a triage algorithm that consists of 52 flowcharts, covering patients' chief signs and symptoms such as "Headache", "Shortness of breath" and "Wounds". Each flowchart in turn consists of additional signs and symptoms named discriminators, such as "Airway compromise", "Severe pain" or "Persistent vomiting", which are ranked by priority. General discriminators appear throughout the different charts while specific discriminators apply to small groups of presentations. Triage nurses select for each patient the most appropriate flowchart and consequently gather information on the discriminators from top to bottom. Selection of a discriminator allocates the patient to the related urgency category, ranging from "immediate" (0 minutes maximum waiting time) to "non-urgent" (240 minutes maximum waiting time). A discriminator will lead to the same urgency level, regardless of the flowchart used, increasing the ease of use and the interrater reliability.

In all three hospitals, trained nurses perform triage with a computerized triage

application. Both Erasmus MC and Maasstad Hospital use the official Dutch translation of the second edition of the MTS.[12, 58] In the Erasmus MC, some specific modifications for children are implemented based on previous research.[35] The main difference includes a modification for children with fever. The Hospital Fernando Fonseca uses the official Portuguese translation of the second edition of the MTS which includes already some of the modifications implemented in the third edition of the MTS.[59] These differences consist of adaptations for children with fever, and the addition of a small number of extra discriminators. Details on the different versions of the MTS used in the study are provided in the supporting information (Table S1).

3- category reference standard

Before the study started, a reference standard as proxy for patients' true urgency was developed. We defined several requirements for our reference standard. It had to be a good proxy for patient urgency, independent of triage, be applied to individual patients with a wide range of problems, contain objective items that could be compared between settings, and identify at least 3 urgency levels to allow for evaluation of modifications.[29] First, we performed a literature review to identify currently used reference standards for triage. None of the reference standards fulfilled all our requirements.[60] Therefore, we composed an expert panel, consisting of a neurologist, a surgeon specialized in traumatology, an internist specialized in intensive care, a cardiologists, an emergency physician and a pediatrician. In an evaluation meeting, the panel discussed the individual reference standards and combined a selection of the most relevant items into a multilevel reference standard.

The final reference standard, as presented in Table 1, consisted of three urgency categories based on a combination of vital signs, treatment at the emergency department and patient disposition. Vital signs were measured at discretion of the nurse, and therefore not all patients had a complete set of vital signs recorded. If vital signs were not documented, they were assumed to be normal, which is in agreement with clinical experience. The low number of vital signs documented in the least urgent patients (e.g. heart rate was measured in 74% of patients in reference category 1 versus 36% in reference category 3) and the co-occurrence of missing vital signs in the same patients made it impossible to perform multiple imputation to handle the missing data. However, these findings also support our assumption that patients with missing vital signs in the absence of any other positive reference standard item are unlikely to be urgent.

Table 1. 3-category reference standard as proxy for true patient urgency

Category	Corresponding MTS category	Maximum waiting time (minutes)	Items adults	Items children
R1	Immediate and Very urgent	0- 10	Abnormal vital signs as defined by a modified early warning score ≥5 [61]	Abnormal vital signs according to a previously used reference standard,[31] based on the pediatric risk of mortality score (PRISM III) [40]
			Level of consciousness reacting to pain or unresponsive	Level of consciousness reacting to pain or unresponsive
			Mortality at the ED, ICU or high care admission*	Mortality at the ED or ICU admission
			Emergency surgery <4hours after arrival, including cardiac catheterization and endovascular aortic repair procedures	
R2	Urgent	60	- IV medication, fluids or nebulizers at the ED	- IV medication, fluids or nebulizers at the ED
			- Hospitalization	- Hospitalization
R3	Standard and Non-urgent	120 - 240	None of the above	None of the above

*Patients at hospital Fernando Fonseca do not have information on high care admission or emergency surgery available

Data collection

Data on patient characteristics, triage, vital signs, resource utilization, admission to hospital, and follow-up are routinely documented in all hospitals and were automatically extracted from the electronic hospital information systems. Trained medical students entered data that was only available on paper emergency department forms in a separate database, blinded to MTS urgency, using SPSS Data entry version 4.0.

Data analysis

First, validity of the MTS high urgency categories ("immediate" and "very urgent") was assessed for the identification of patients requiring ICU admission, including the patients that died at the ED. We included ICU admission and death as a separate reference standard, because it has a strong correlation with patient urgency and is relatively independent of the clinical setting.

Second, for each individual patient, a reference standard category was determined, based on the 3-category reference standard. We assessed validity of the MTS by comparing the allocated MTS urgency category with the reference urgency category.

Validity was assessed by the proportion of correctly triaged, undertriaged and overtriaged patients and by the different diagnostic performance measures sensitivity,

specificity, positive and negative likelihood ratio and diagnostic odds ratio. Undertriage was defined as the proportion of patients who were allocated to a lower MTS urgency category than the reference category and overtriage as the proportion of patients allocated to a higher MTS urgency category than the reference category. To calculate the diagnostic performance measures, we dichotomized the MTS and the reference standard into high (MTS category "immediate" and "very urgent"; reference category 1) and low urgency (MTS category "urgent", "standard" and "non-urgent"; reference category 2 and 3). Sensitivity analyses were performed to assess the impact on MTS performance of the modifications for children with fever that were adopted in the Erasmus MC and Hospital Fernando Fonseca.[19,20] We did not assess the effect of other modifications because these were all together only applied in 1.9% of patients.

The MTS was validated for different subgroups based on age, and we determined five clinically relevant age groups: <1 year, 1 to 16 years, 16 to 65 years, 65 to 80 years and ≥80 years. Finally, we assessed validity of the most commonly used flowcharts and general discriminators in adult patients. Discriminators were grouped into the hemorrhage, consciousness and temperature discriminators.[14] We compared performance of these flowcharts and discriminators in adult patients with performance in the subgroup of patients aged 65 and older.

Analyses were performed using SPSS software (version 20.0). Diagnostic performance measures with 95% confidence intervals were calculated with the VassarStats website (http://statline.cbs.nl/statweb).

RESULTS

Characteristics of study subjects

During the study period 306,090 patients attended the emergency department of one of the three hospitals. After the exclusion of patients with incomplete information on triage or reference standard items, 288,663 patients (94.3%) were available for analysis: 25,583 from the Erasmus MC, 32,532 from the Maasstad Hospital and 230,548 from Hospital Fernando Fonseca (Supporting information, Fig S2). The Erasmus MC has a relatively high percentage of missing MTS urgency, which can be explained by the absence of triage nurses during night shifts at the start of the study. Since these missing values are expected to be at random, we used a complete case approach.

Hospital Fernando Fonseca has the largest caseload while the two Dutch hospitals have the most severe case-mix with a larger percentage of hospital and ICU admitted patients. Further characteristics of the study populations are presented in Table 2.

Table 2. Characteristics of the study population

	Erasmus MC n=25,583	Maasstad n=32,532	Fernando Fonseca n=230,548
Age categories, n (%)			
0-16 years	6185 (24.2)	7032 (21.6)	52,843 (22.9)
16-65 years	15,980 (62.5)	18,226 (56.0)	127,562 (55.3)
≥65 years	3418 (13.4)	7274 (22.4)	50,143 (21.7)
Gender, n (%)			
Male	14,611 (57.1)	16,600 (51.0)	99,406 (43.1)
Female	10,972 (42.9)	15,932 (49.0)	131,142 (56.9)
Presenting problem, n (%)			
Cardiac	1780 (7.0)	993 (3.1)	14,185 (6.2)
Dermatological	2960 (11.6)	3969 (12.2)	22,251 (9.7)
Ear, Nose and Throat	796 (3.1)	475 (1.5)	20,236 (8.8)
Gastrointestinal	3109 (12.2)	4681 (14.4)	29,101 (12.6)
Neurologic or psychiatric	2644 (10.3)	1769 (5.4)	16,217 (7.0)
Respiratory	1631 (6.4)	3079 (9.5)	21,955 (9.5)
Trauma or muscular	7536 (29.5)	11,689 (35.9)	53,711 (23.3)
General malaise	3304 (12.9)	3463 (10.6)	16,869 (7.3)
Uro- or gynaecological	752 (2.9)	620 (1.9)	18,422 (8.0)
Other or unknown	1071 (4.2)	1794 (5.5)	17,601 (7.6)
MTS urgency, n (%)			
Immediate	432 (1.7)	208 (0.6)	1365 (0.6)
Very urgent	2425 (9.5)	5075 (15.6)	37,502 (16.3)
Urgent	11,516 (45.0)	16,811 (51.7)	76,777 (33.3)
Standard	11,016 (43.1)	10,332 (31.8)	109,956 (47.7)
Non-urgent	194 (0.8)	106 (0.3)	4948 (2.1)
Disposition, n (%)			
Hospital admission	6914 (27.0)	9472 (29.1)	26,832 (11.6)
ICU admission	438 (1.7)	245 (0.8)	461 (0.2)
Mortality at the ED	43 (0.2)	32 (0.1)	74 (<0.1)

Overall validity of the MTS

Sensitivity of the MTS to identify patients that died at the ED or were in need of ICU admission ranged from 0.80 to 0.86 in adults and 0.66 to 0.91 in children. Specificity ranged from 0.84 to 0.91 in adults and 0.82 to 0.87 in children (Table 3). This performance varied considerably between the different settings. Overall performance as indicated by the diagnostic odds ratio was lower in children than in adults except in the Maasstad hospital. However, the absolute number of children admitted to ICU in this hospital was very small.

Table 3. Diagnostic performance of the MTS for the identification of patients who died at the emergency department or required ICU admission

	Erasmus MC		Maasstad		Fernando Fonseca	
	<16 years n=6185	**≥16 years n=19,398**	**<16 years n=7032**	**≥16 years n=25,500**	**<16 years n=52,843**	**≥16 years n=177,705**
Total ICU admissions, n (%)	148 (2.4)	333 (1.7%)	11 (0.2%)	266 (1.0%)	132 (0.2%)	403 (0.2%)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.66 (0.58-0.73)	0.80 (0.76-0.84)	0.91 (0.62-0.98)	0.86 (0.81-0.90)	0.77 (0.69-0.83)	0.84 (0.80-0.87)
Specificity	0.87 (0.86-0.88)	0.91 (0.91-0.92)	0.83 (0.82-0.84)	0.85 (0.84-0.85)	0.82 (0.82-0.83)	0.84 (0.83-0.84)
Positive Likelihood Ratio	4.92 (4.30-5.62)	9.10 (8.48-9.75)	5.26 (4.34-6.39)	5.67 (5.36-6.00)	4.33 (3.94-4.77)	5.12 (4.90-5.35)
Negative Likelihood Ratio	0.40 (0.32-0.50)	0.21 (0.17-0.27)	0.11 (0.02-0.71)	0.16 (0.12-0.22)	0.29 (0.21-0.39)	0.19 (0.15-0.24)
Diagnostic Odds Ratio	12.4 (8.7-17.5)	42.5 (32.2-55.9)	47.9 (6.1-374.4)	34.6 (24.4-49.0)	15.2 (10.2-22.7)	27.0 (20.6-35.2)

When using the predefined 3-category reference classification, the MTS agreed with the reference standard in 61.6% of adult patients in the Erasmus MC, 49.7% in Maasstad Hospital and 51.7% in the Fernando Fonseca Hospital. In children, these percentages were 50.2%, 46.0% and 59.6% respectively. Overtriage was much more common than undertriage with percentages ranging from 26.9% to 44.0% in adults and 36.9% to 50.3% in children. Undertriage was present in 6.2% to 14.1% of adults and 3.5% to 5.8% of children.

Sensitivity to detect high urgent patients was moderate in the two Dutch hospitals and good in the Fernando Fonseca while specificity was good in all three hospitals. A summary of all diagnostic performance measures are presented in Table 4. The numbers of correct, over- and undertriage per MTS category are presented in the Supporting information (Tables S3)

Sensitivity analyses showed that the modifications for children with fever improved performance in both settings. Without the modifications, the MTS would have had a slightly higher sensitivity at the cost of a lower specificity in the Erasmus MC, while in the hospital Fernando Fonseca sensitivity would be similar with a lower specificity (Supporting information, Tables S4).

Table 4. Diagnostic performance of the MTS, as determined by the 3-category reference standard

	Erasmus MC		Maasstad		Fernando Fonseca	
	<16 years n=6185	≥16 years n=19,398	<16 years n=7032	≥16 years n=25,500	<16 years n=52,843	≥16 years n=177,705
Absolute classification (%)						
Correct triage	3104 (50.2)	11,940 (61.6)	3232 (46.0)	12,685 (49.7)	31,506 (59.6)	91,796 (51.7)
Overtriage	2722 (44.0)	5221 (26.9)	3534 (50.3)	11,228 (44.0)	19,487 (36.9)	60,928 (34.3)
Undertriage	359 (5.8)	2237 (11.5)	266 (3.8)	1587 (6.2)	1850 (3.5)	24,981 (14.1)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.65 (0.61-0.70)	0.47 (0.44-0.49)	0.66 (0.57-0.74)	0.72 (0.70-0.75)	0.83 (0.79-0.87)	0.87 (0.85-0.90)
Specificity	0.89 (0.88-0.90)	0.94 (0.94-0.94)	0.83 (0.83-0.84)	0.87 (0.87-0.87)	0.83 (0.82-0.83)	0.84 (0.84-0.84)
Positive likelihood ratio	6.12 (5.54-6.78)	7.66 (7.11-8.26)	3.99 (3.47-4.59)	5.59 (5.33-5.86)	4.79 (4.55-5.05)	5.36 (5.20-5.52)
Negative likelihood ratio	0.39 (0.34-0.44)	0.57 (0.55-0.59)	0.41 (0.32-0.52)	0.32 (0.29-0.35)	0.20 (0.16-0.26)	0.15 (0.13-0.18)
Diagnostic Odds Ratio	15.8 (12.8-19.6)	13.5 (12.1-15.0)	9.8 (6.7-14.5)	17.7 (15.5-20.1)	23.8 (17.7-32.0)	35.3 (28.4-43.9)

Performance in different age groups

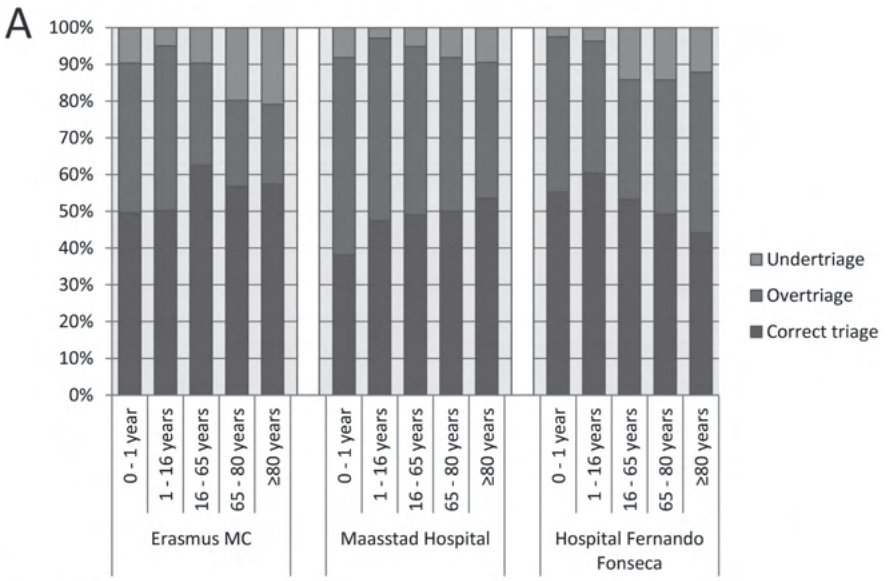
Performance of the MTS in different age groups showed a large variation between settings (Fig 1; Supporting information, Tables S5). Overall, the diagnostic odds ratio was lower in elderly patients, aged 65 or older, when compared to the group of adults aged 16 to 65 and this was more prominent in the patients above the age of 80. While in all three hospitals specificity was lower in the older age groups, sensitivities varied when compared to adult patients.

Children had lower diagnostic odds ratios than the adult groups, except in the Erasmus MC. More specifically, specificity was lower in children compared to adults, but sensitivities varied compared to the adult reference group. There was no clear trend towards a decreased performance of the MTS in the youngest children.

Performance of different flowcharts and discriminators

In adults, the most commonly used flowcharts in the three hospitals were "Limb problems", "Unwell adult", "Abdominal pain in adults", "Chest pain", "Shortness of breath in adults" and "Headache", together accounting for 39% of adult patients. The general discriminators most often used were the consciousness and temperature discriminators, together accounting for 3.2% of adult patients. In hospital Fernando Fonseca, relatively few patients were triaged as high urgent and therefore performance could not be assessed for all flowcharts and discriminators.

Fig 1. Performance of the MTS in different age groups.



B

	Diagnostic Odds Ratio		
	Erasmus MC	Maasstad Hospital	Fernando Fonseca
0-1 years	14.4 (9.1 to 22.9)	16.4 (4.9 to 55.3)	10.8 (6.3 to 18.7)
1-16 years	16.1 (12.6 to 20.4)	9.6 (6.3 to 14.7)	29.4 (20.7 to 41.7)
16-64 years	15.8 (13.8 to 18.0)	24.1 (20.3 to 28.6)	38.2 (28.3 to 51.5)
65-80 years	7.7 (6.1 to 9.8)	11.3 (8.8 to 14.6)	30.5 (19.7 to 47.1)
≥ 80 years	7.5 (5.1 to 10.9)	8.0 (5.9 to 11.0)	20.1 (12.5 to 32.5)

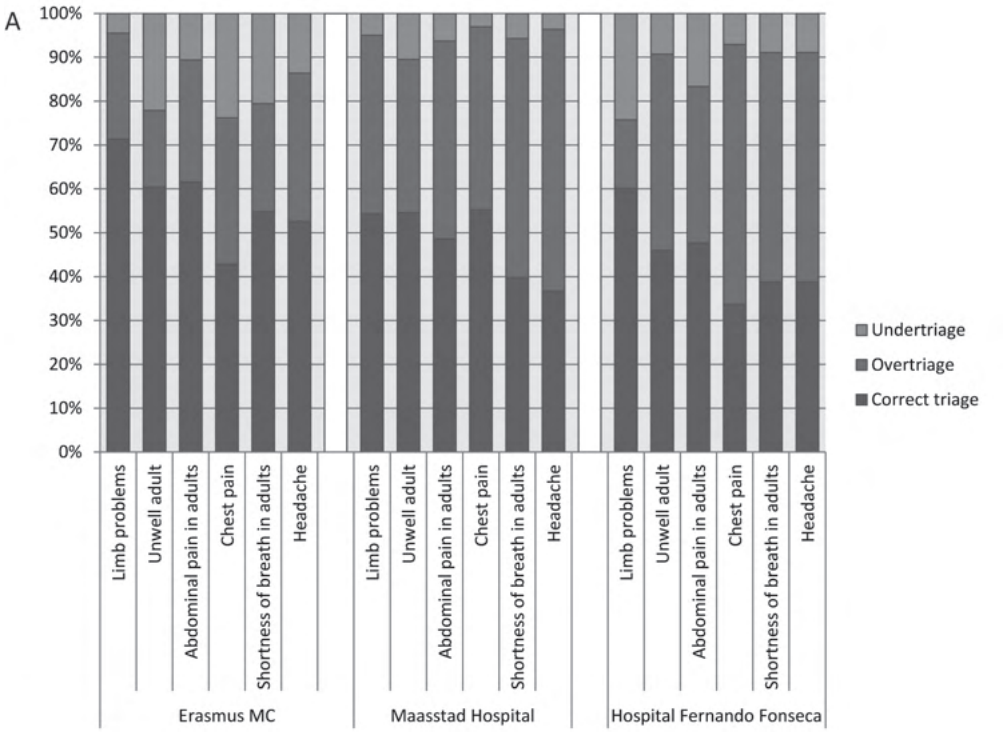
A) percentages under-, over-, and correct triage; B) diagnostic odds ratio's

Overall, there was a large variation between settings in performance of the flowcharts and discriminators (Figs 2 and 3; Supporting information, Tables S6 and S7) although performance in general was best in the hospital Fernando Fonseca and poorest in the Erasmus MC. In particular, sensitivities of the flowcharts and discriminators were very low. The flowcharts "Limb problems", "Unwell adult", "Abdominal pain in adults" and "Chest pain" even had sensitivities below the value of 0.5. The temperature discriminators had in all settings low sensitivities with high specificities, while consciousness discriminators had better sensitivities with moderate specificities.

Overall, there was a lower performance of the most commonly used flowcharts and discriminators in the elderly patients.

Fig 2. Performance of most commonly used MTS flowcharts

A



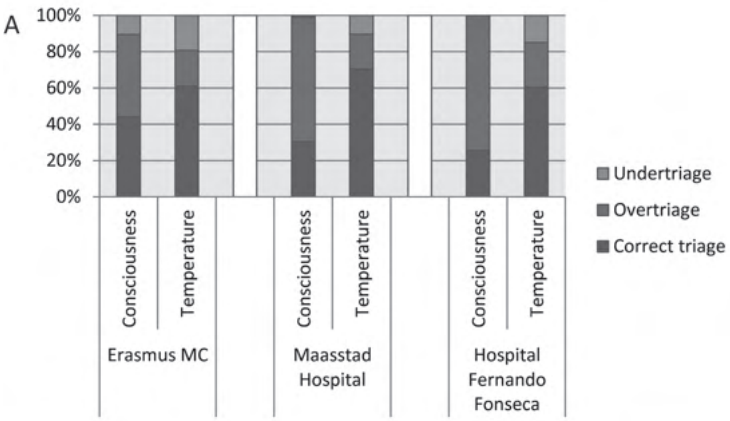
B

	Diagnostic Odds Ratio					
	Erasmus MC		Maasstad Hospital		Fernando Fonseca	
	All adults	Elderly	All adults	Elderly	All adults	Elderly
Limb problems	14.7 (7.0 to 30.6)	5.5 (1.4 to 21.3)	15.5 (9.2 to 26.0)	3.7 (1.6 to 8.5)	*	*
Unwell adult	13.6 (9.7 to 19.0)	8.2 (4.8 to 14.2)	7.5 (5.5 to 10.4)	7.6 (4.9 to 11.8)	29.1 (17.4 to 48.8)	21.6 (11.3 to 41.2)
Abdominal pain in adults	8.2 (4.0 to 16.8)	*	14.8 (9.0 to 24.6)	11.4 (5.9 to 22.1)	*	*
Chest pain	4.7 (3.6 to 6.2)	3.3 (2.0 to 5.5)	22.9 (15.0 to 35.1)	12.0 (4.5 to 31.9)	14.3 (9.1 to 22.5)	10.2 (4.7 to 22.2)
Shortness of breath in adults	6.7 (4.8 to 9.4)	4.1 (2.5 to 6.6)	7.4 (4.9 to 11.0)	5.7 (3.4 to 9.6)	12.6 (6.7 to 23.5)	11.4 (4.9 to 26.2)
Headache	9.8 (6.8 to 14.1)	6.9 (3.4 to 14.0)	8.0 (2.9 to 22.1)	6.4 (1.8 to 22.9)	19.4 (6.4 to 59.1)	*

**≤10 high urgent patients available for analysis*

A) Percentages under-, over-, and correct triage; B) Diagnostic odds ratio's

Fig 3. Performance of most commonly used MTS discriminators



B

	Diagnostic Odds Ratio					
	Erasmus MC		Maasstad Hospital		Fernando Fonseca	
	All adults	Elderly	All adults	Elderly	All adults	Elderly
Consciousness**	6.2 (4.2 to 9.1)	2.4 (1.2 to 4.9)	13.1 (5.5 to 30.9)	7.2 (2.7 to 19.2)	49.0 (15.6 to 153.5)	27.4 (6.8 to 111.4)
Temperature***	5.8 (3.2 to 10.4)	5.1 (1.5 to 17.4)	3.8 (1.7 to 8.2)	2.8 (1.0 to 7.9)	35.7 (10.9 to 116.7)	*

**≤10 high urgent patients available for analysis*

*** Consciousness discriminators: "Currently fitting", "Altered consciousness level", "History of unconsciousness"*

**** Temperature discriminators: "Cold", "Very hot adult", "Hot adult", "Warmth"*

A) Percentages under-, over-, and correct triage; B) Diagnostic odds ratio's

DISCUSSION

This multicenter observational study demonstrates that validity of the MTS for emergency department triage is moderate to good. When compared to a predefined, 3-category reference standard, sensitivity was 0.47 to 0.87 and specificity 0.84 to 0.94 for the triage of adult patients while sensitivity was 0.65 to 0.83 and specificity 0.83 to 0.89 for the triage of children. In all three hospitals, overall validity as determined by the diagnostic odds ratio was lowest in the youngest and oldest patients. One of the most remarkable findings was the high variability in performance of the MTS between the different emergency departments.

Previous studies have assessed performance of the MTS, the majority by evaluating associations between MTS triage category and hospitalization or resource use.[42-44, 54] Our study shows that specificity of the MTS when compared to a 3-category reference

standard was very good, but sensitivity was moderate in two of the three hospitals. A low sensitivity indicates that high urgent patients are being "missed" by the triage system, which leads to longer waiting times for these patients and poses them at risk for adverse outcomes due to harm by delay in treatment. In our study, validity of the MTS for the most urgent patients, i.e. those requiring ICU admission, was better, but in absolute numbers the MTS still classifies 14 to 20% of adults and 9 to 34% of children in need of ICU admission as low urgent. These results indicate that improvement of the MTS is still needed.

Importantly, we found that performance of the MTS was lowest in the young and elderly patients. To our knowledge, this is the first study that assesses performance of the MTS for specific age groups. Only one study on the triage of patients with acute myocardial infarction specifically looked at age and found that patients above the age of 70 were less often correctly triaged as high urgent by the MTS.[56] There is also some evidence from the Emergency Severity Index and several trauma triage systems that elderly patients are at risk of undertriage.[62-65] Previous modifications targeted at children have been shown to improve validity of the MTS and were consequently partially implemented in the most recent MTS edition.[14, 35] Likewise, modifications aimed at elderly might be a promising way to improve triage for this patient group.

Our results show a remarkable variation between the three hospitals and we believe this can be explained by several factors. First, the differences in patient population attending the different emergency departments is likely to influence the validity of triage systems. It is well known that population characteristics, including demographic features, disease severity and disease prevalence influence the performance of diagnostic tests.[66] In the case of triage, it can be expected that increased patient complexity contributes to lower performance of a triage system because patients with rare disorders or multiple comorbid conditions may be more difficult to triage.[34] This could explain the lower performance of the MTS in the Erasmus MC, which is a tertiary hospital receiving relatively large numbers of complex patients. It is also possible, that disease prevalence plays a role and nurses apply triage criteria more strictly in settings with a lower prevalence of urgent patients, compared to settings with a higher prevalence. Secondly, some of the differences in performance of the MTS can be explained by the differences in availability of the reference standard items. The hospital Fernando Fonseca did not record information on high care admission, and emergency surgery, so misclassification of the outcome in some of the high urgent patients might have led to an overestimation of the validity of the MTS in this hospital. Moreover, it is possible that potential differences in clinical practice and different indications for reference standard items such as hospitalization and intravenous medication can explain some of the variability in the results. Nevertheless, this is probably not the entire explanation, as we also observed differences when using ICU admission as the reference standard, while we consider indications for ICU admission approximately

similar in the three settings. Modifications of the MTS in the different hospitals may only have influenced the results marginally. Differences in MTS version between the hospitals were minor in adults, and in children only had a moderate impact on sensitivity and specificity in the Erasmus MC.

Moreover, we do not believe that differences in application of the MTS by the triage nurses have caused this large variation in results. Nurses in all three hospitals receive formal training in the MTS before they are allowed to triage patients, and previous studies, performed in different settings, showed that interrater reliability of the MTS was moderate to good.[42, 45, 67, 68] Even though a combination of patient and hospital related factors might contribute to the variability in performance of the MTS, this is simply a reflection of clinical practice. Variability in emergency department size, population and practices throughout the world simply exist, and triage needs to be conducted in any of these circumstances. Future multicenter studies should therefore focus on unravelling the factors that explain variability in triage performance between clinical settings. Consequently, it would be important to determine whether and how triage systems can be adapted to the local circumstances. Until then, our study indicates that the results of single-center studies evaluating a triage system should be interpreted with caution.

Our study has several strengths: it is based on a large cohort of almost 300,000 emergency department visits and it includes data from three different clinical settings, which increases generalizability. Moreover, we had less than 6% missing data on triage and reference standard items and we therefore believe risk of selection bias is low.

We assessed validity of the MTS by a 3-category reference standard as a proxy for true patient urgency, developed by a panel with expertise in the field of emergency care, and consisting of items undoubtedly related to patient urgency. In the absence of a golden standard for the evaluation of triage systems, the combination of multiple items to construct a reference standard is a valid approach.[69] Previous studies have evaluate triage systems with several single outcome measures such as hospitalization or resource use, which can merely be used to display trends in a certain outcome over different urgency categories. None of these individual items is able to perfectly distinguish the high from the low urgent patients. The combination of different items is a more precise way to describe true patient urgency and enables the evaluation of modifications to different triage categories. Still, our reference standard is a proxy of true patient urgency and therefore our results represent an estimation of the validity of the MTS.

A limitation of our study is that one of the hospitals did not have information on high care admission or emergency surgery, two items in the reference standard. This difference makes the interpretation of the results more difficult. Therefore, we also assessed ICU

admission as another reference standard, which was collected in a similar way in all three settings. Moreover, due to the observational design of the study, based on routine data, we have to accept the occurrence of missing data, particularly in the documentation of vital signs. We assumed vital signs that were not measured to be normal, and although this is in line with clinical experience, and may be true for the majority of patients, we cannot exclude that we misclassified a small proportion of patients with abnormal vital signs that were not recorded.

Although our study has a large sample size, we still had insufficient numbers of high urgent patients available to derive and validate modifications for specific patient subgroups. This was also the case for the assessment of validity for specific flowcharts and discriminators, indicating that future studies should include at least a substantial number of high urgent patients per subgroup, specific flowchart or discriminator.

CONCLUSION

This study shows that validity of the MTS is moderate to good, with poorer performance in the most vulnerable patient populations: the young and elderly. Moreover, the study reveals that it matters where you validate a triage system, since results are highly variable between the different clinical settings. Due to the large variability between the emergency departments, we could not propose modifications to improve the MTS. Future research should therefore be restricted to large multicenter studies, and conducted in diverse hospitals. This way, potential modifications to improve the MTS can be defined, and results can be generalized or adapted to different clinical settings.

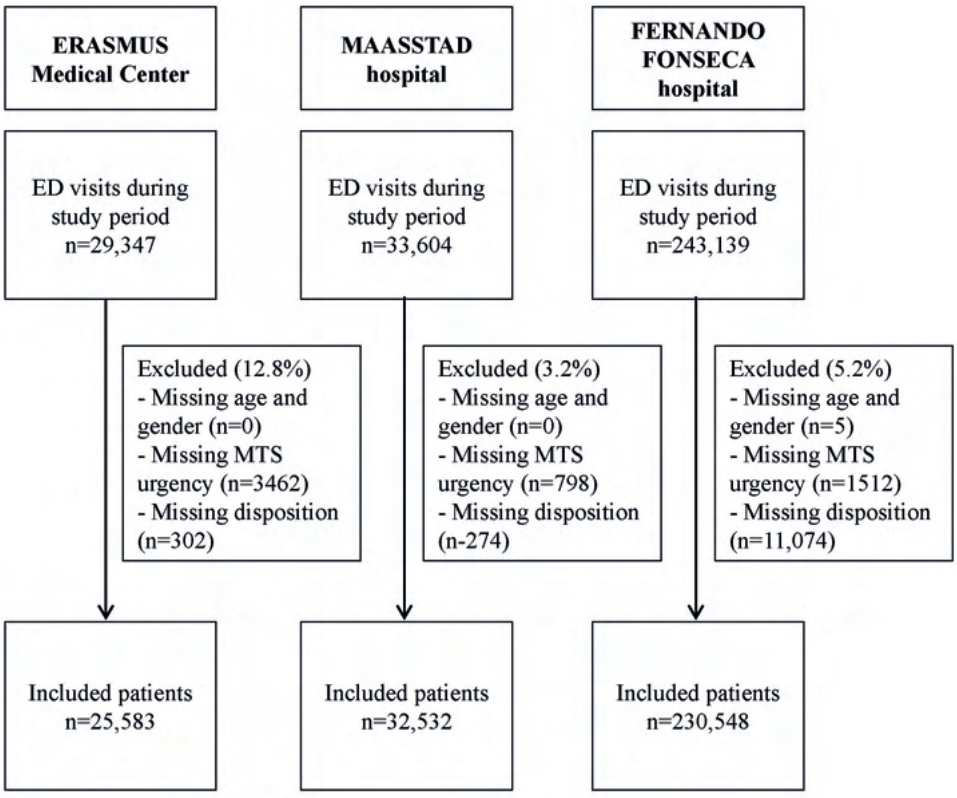
ACKNOWLEDGMENTS

We thank Wouter van Rijsinge of Erasmus MC and Peter Keller of Maasstad Hospital for their technical support during the data extraction. We thank Jos de Bruijn and Nicole van Geesink of Erasmus MC for their support of the study. We thank the members of the expert panel that developed the reference standard for their expertise. Also, we would like to thank all ED nurses for making the study possible.

SUPPORTING INFORMATION

Available as online web appendix on the website of PLOS One:
- S1 Table. Differences between the MTS versions used in the three settings
- S2 Table. Raw data
- S2 File. Sensitivity analysis of MTS performance, comparing validity of the MTS with and without modifications for children with fever

S1 Fig. Flow diagram of the study population



S1 File. Numbers of correct, over- and undertriage per MTS category

Table A. Erasmus MC, <16 years

	Reference category 1	Reference category 2	Reference category 3	Total
Immediate	99	16	11	126
Very urgent	189	268	319	776
Urgent	113	599	2108	2820
Standard	37	193	2061	2291
Non-urgent	2	14	156	172
Total	440	1090	4655	6185

correct triage overtriage undertriage

Table B. Maasstad, <16 years

	Reference category 1	Reference category 2	Reference category 3	Total
Immediate	4	9	2	15
Very urgent	74	530	604	1208
Urgent	30	661	2389	3080
Standard	10	224	2475	2709
Non-urgent	0	2	18	20
Total	118	1426	5488	7032

correct triage overtriage undertriage

Table C. Fernando Fonseca, <16 years

	Reference category 1	Reference category 2	Reference category 3	Total
Immediate	21	135	80	236
Very urgent	245	1488	7437	9170
Urgent	28	1516	10347	11891
Standard	24	1778	29072	30874
Non-urgent	1	19	652	672
Total	319	4936	47588	52843

correct triage overtriage undertriage

Table D. Erasmus MC, ≥16 years

	Reference category 1	Reference category 2	Reference category 3	Total
Immediate	256	43	7	306
Very urgent	635	779	235	1649
Urgent	890	3649	4157	8696
Standard	130	1212	7383	8725
Non-urgent	0	5	17	22
Total	1911	5688	11799	19398

correct triage overtriage undertriage

Table E. Maasstad, ≥16 years

	Reference category 1	Reference category 2	Reference category 3	Total
Immediate	143	40	10	193
Very urgent	773	2020	1074	3867
Urgent	313	5334	8084	13731
Standard	35	1227	6361	7623
Non-urgent	0	12	74	86
Total	1264	8633	15603	25500

correct triage overtriage undertriage

Table F. Fernando Fonseca, ≥16 years

	Reference category 1	Reference category 2	Reference category 3	Total
Immediate	94	939	96	1129
Very urgent	538	19654	8140	28332
Urgent	76	32711	32099	64886
Standard	15	24105	54962	79082
Non-urgent	1	784	3491	4276
Total	724	78193	98788	177705

correct triage overtriage undertriage

S3 File. Diagnostic performance of the MTS for different age groups, as determined by the 3-category reference standard

Table A. Erasmus MC

Erasmus MC					
	0-1 years n=1107	**1-16 years n=5078**	**16-65 years n=15,980**	**65-80 years n=2598**	**≥ 80 years n=820**
Absolute classification (%)					
Correct triage	549 (49.6)	2555 (50.3)	9994 (62.5)	1475 (56.8)	471 (57.4)
Overtriage	451 (40.7)	2271 (44.7)	4435 (27.8)	608 (23.4)	178 (21.7)
Undertriage	107 (9.7)	252 (5.0)	1551 (9.7)	515 (19.8)	171 (20.9)
Diagnostic accuracy (95% confidence interval)					
Sensitivity	0.69 (0.60 to 0.77)	0.64 (0.59 to 0.69)	0.48 (0.45 to 0.50)	0.43 (0.39 to 0.48)	0.49 (0.42 to 0.56)
Specificity	0.86 (0.84 to 0.88)	0.90 (0.89 to 0.91)	0.95 (0.94 to 0.95)	0.91 (0.90 to 0.92)	0.89 (0.86 to 0.91)
Positive Likelihood Ratio	5.13 (4.18 to 6.28)	6.38 (5.68 to 7.16)	8.75 (8.01 to 9.57)	4.82 (4.07 to 5.71)	4.30 (3.30 to 5.59)
Negative Likelihood Ratio	0.35 (0.26 to 0.48)	0.40 (0.34 to 0.46)	0.55 (0.53 to 0.58)	0.62 (0.58 to 0.68)	0.58 (0.50 to 0.67)
Diagnostic Odds Ratio	14.4 (9.1 to 22.9)	16.1 (12.6 to 20.4)	15.8 (13.8 to 18.0)	7.7 (6.1 to 9.8)	7.5 (5.1 to 10.9)

Table B. Maasstad

Maasstad					
	0-1 years n=1212	**1-16 years n=5820**	**16-65 years n=18,226**	**65 - 80 years n=4494**	**≥ 80 years n=2780**
Absolute classification (%)					
Correct triage	463 (38.2)	2769 (47.6)	8945 (49.1)	2249 (50.0)	1491 (53.6)
Overtriage	650 (53.6)	2884 (49.6)	8327 (45.7)	1875(41.7)	1026 (36.9)
Undertriage	99 (8.2)	167 (2.9)	954 (5.2)	370 (8.2)	263 (9.5)
Diagnostic accuracy (95% confidence interval)					
Sensitivity	0.88 (0.69 to 0.96)	0.61 (0.51 to 0.70)	0.73 (0.70 to 0.77)	0.74 (0.69 to 0.78)	0.67 (0.60 to 0.73)
Specificity	0.70 (0.67 to 0.73)	0.86 (0.85 to 0.87)	0.90 (0.89 to 0.90)	0.80 (0.79 to 0.81)	0.80 (0.78 to 0.81)
Positive Likelihood Ratio	2.92 (2.45 to 3.48)	4.40 (3.69 to 5.24)	7.13 (6.70 to 7.58)	3.73 (3.41 to 4.08)	3.32 (2.93 to 3.76)
Negative Likelihood Ratio	0.18 (0.06 to 0.52)	0.46 (0.36 to 0.59)	0.30 (0.26 to 0.33)	0.33 (0.28 to 0.39)	0.41 (0.34 to 0.51)
Diagnostic Odds Ratio	16.4 (4.8 to 55.2)	9.6 (6.3 to 14.7)	24.1 (20.3 to 28.6)	11.3 (8.8 to 14.6)	8.0 (5.9 to 11.0)

Table C. Fernando Fonseca

Fernando Fonseca					
	0-1 years n=8185	**1-16 years n=44,658**	**16-65 years n=127,562**	**65 - 80 years n=32,689**	**≥ 80 years n=17,454**
Absolute classification (%)					
Correct triage	4518 (55.2)	26,988 (60.4)	68,012 (53.3)	16,071 (49.2)	7713 (44.2)
Overtriage	3465 (42.3)	16,022 (35.9)	41,376 (32.4)	11,943 (36.5)	7609 (43.6)
Undertriage	202 (2.5)	1648 (3.7)	18,174 (14.2)	4675 (14.3)	2132 (12.2)
Diagnostic accuracy (95% confidence interval)					
Sensitivity	0.81 (0.71 to 0.88)	0.84 (0.79 to 0.88)	0.86 (0.82 to 0.89)	0.89 (0.84 to 0.92)	0.89 (0.83 to 0.93)
Specificity	0.72 (0.71 to 0.73)	0.85 (0.84 to 0.85)	0.86 (0.86 to 0.87)	0.79 (0.79 to 0.80)	0.72 (0.71 to 0.72)
Positive LR	2.89 (2.59 to 3.23)	5.45 (5.13 to 5.78)	6.31 (6.04 to 6.60)	4.32 (4.10 to 4.56)	3.14 (2.96 to 3.33)
Negative LR	0.27 (0.17 to 0.42)	0.19 (0.14 to 0.25)	0.17 (0.13 to 0.21)	0.14 (0.10 to 0.21)	0.16 (0.10 to 0.24)
DOR	10.8 (6.3 to 18.7)	29.4 (20.7 to 41.7)	38.2 (28.3 to 51.5)	30.5 (19.7 to 47.1)	20.1 (12.5 to 32.5)

S4 File. Diagnostic performance of the MTS for the most commonly used MTS flowchart, as determined by the 3-category reference standard

Table A. Flowchart Limb problems

	Erasmus MC		Maasstad		Fernando Fonseca	
	All adults n=4021	**Elderly n=415**	**All adults n=6829**	**Elderly n=1431**	**All adults n=24,723**	**Elderly n=6504**
High urgent patients, n (%)	80 (2.0)	16 (3.9)	144 (2.1)	63 (4.4)	528 (2.1)	183 (2.8)
Absolute classification (%)						
Correct triage	2870 (71.4)	253 (61.0)	3705 (54.3)	722 (50.5)	14,871 (60.2)	3551 (54.6)
Overtriage	972 (24.2)	116 (28.0)	2782 (40.7)	584 (40.8)	3845 (15.6)	1061 (16.3)
Undertriage	179 (4.5)	46 (11.1)	342 (5.0)	125 (8.7)	6007 (24.3)	1892 (29.1)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.21 (0.12-0.34)	0.16 (0.06-0.38)	0.22 (0.15-0.32)	0.13 (0.07-0.25)	*	*
Specificity	0.98 (0.98-0.99)	0.97 (0.94-0.98)	0.98 (0.98-0.98)	0.96 (0.95-0.97)		
Positive LR	11.82 (6.50-21.51)	4.81 (1.50-15.47)	12.23 (8.07-18.54)	3.31 (1.59-6.92)		
Negative LR	0.81 (0.70-0.93)	0.87 (0.72-1.06)	0.79 (0.71-0.88)	0.90 (0.81-1.01)		
DOR	14.7 (7.0-30.6)	5.5 (1.4-21.3)	15.5 (9.2-25.9)	3.7 (1.6-8.5)		

*≤10 high urgent patients available for analysis

Table B. Flowchart Unwell adult

	Erasmus MC		Maasstad		Fernando Fonseca	
	All adults n=2429	**Elderly n=672**	**All adults n=2541**	**Elderly n=1254**	**All adults n=13,272**	**Elderly n=6117**
High urgent patients, n (%)	170 (7.0)	70 (10.4)	463 (18.2)	263 (21.0)	3507 (26.4)	2199 (35.9)
Absolute classification (%)						
Correct triage	1467 (60.4)	388 (57.7)	1387 (54.6)	742 (59.2)	6101 (46.0)	2451 (40.1)
Overtriage	426(17.5)	104 (15.5)	887 (34.9)	357 (28.5)	5934 (44.7)	3120 (51.0)
Undertriage	536 (22.1)	180 (26.8)	267 (10.5)	155 (12.4)	1237 (9.3)	546 (8.9)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.33 (0.28-0.39)	0.36 (0.27-0.47)	0.58 (0.50-0.65)	0.62 (0.52-0.71)	0.91 (0.86-0.94)	0.92 (0.86-0.96)
Specificity	0.96 (0.96-0.97)	0.93 (0.91-0.95)	0.85 (0.83-0.86)	0.82 (0.80-0.85)	0.74 (0.74-0.75)	0.65 (0.64-0.66)
Positive LR	9.42 (7.14-12.42)	5.59 (3.70-8.45)	3.77 (3.22-4.42)	3.53 (2.89-4.31)	3.56 (3.37-3.76)	2.65 (2.49-2.82)
Negative LR	0.69 (0.64-0.75)	0.68 (0.58-0.80)	0.50 (0.42-0.59)	0.46 (0.36-0.60)	0.12 (0.08-0.20)	0.12 (0.07-0.22)
DOR	13.6 (9.7-19.0)	8.2 (4.8-14.2)	7.5 (5.5-10.4)	7.6 (4.9-11.8)	29.1 (17.4-48.8)	21.6 (11.3-41.2)

Table C. Flowchart Abdominal pain in adults

	Erasmus MC		Maasstad		Fernando Fonseca	
	All adults n=2027	**Elderly n=328**	**All adults n=3310**	**Elderly n=656**	**All adults n=13,872**	**Elderly n=3107**
High urgent patients, n (%)	91 (4.5)	21 (6.4%)	327 (9.9)	106 (16.2)	2302 (16.6)	507 (16.3)
Absolute classification (%)						
Correct triage	1248 (61.6)	231 (70.4)	1610 (48.6)	387 (59.0)	6622(47.7)	1593 (51.3)
Overtriage	564 (27.8)	57 (17.4)	1491 (45.0)	207 (31.6)	4938 (35.6)	861 (27.7)
Undertriage	215 (10.6)	40 (12.2)	209 (6.3)	62 (9.5)	2312 (16.7)	653 (21.0)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.26 (0.15-0.40)	*	0.59 (0.47-0.70)	0.63 (0.48-0.76)	*	*
Specificity	0.96 (0.95-0.97)		0.91 (0.90-0.92)	0.87 (0.84-0.90)		
Positive LR	6.34 (3.65-11.03)		6.66 (5.29-8.37)	4.87 (3.58-6.63)		
Negative LR	0.78 (0.65-0.92)		0.45 (0.34-0.60)	0.43 (0.29-0.63)		
DOR	8.2 (4.0-16.8)		14.8 (8.9-24.6)	11.4 (5.9-22.1)		

*≤10 high urgent patients available for analysis

Table D. Flowchart Chest pain

	Erasmus MC		Maasstad		Fernando Fonseca	
	All adults n=1349	**Elderly n=373**	**All adults n=927**	**Elderly n=220**	**All adults n=12,403**	**Elderly n=3968**
High urgent patients, n (%)	285 (21.1)	95 (25.5)	608 (65.6)	175 (79.5)	4781 (38.5)	1982 (49.9)
Absolute classification (%)						
Correct triage	582 (43.1)	159 (42.6)	512 (55.2)	129 (58.6)	4182 (33.7)	1122 (28.3)
Overtriage	452 (33.5)	89 (23.9)	386 (41.6)	86 (39.1)	7343 (59.2)	2631 (66.3)
Undertriage	315 (23.4)	125 (33.5)	29 (3.1)	5 (2.3)	878 (7.1)	215 (5.4)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.39 (0.34-0.43)	0.37 (0.30-0.44)	0.94 (0.91-0.96)	0.95 (0.90-0.98)	0.90 (0.85-0.93)	0.91 (0.82-0.95)
Specificity	0.88 (0.86-0.90)	0.85 (0.80-0.90)	0.60 (0.56-0.65)	0.36 (0.28-0.45)	0.62 (0.61-0.63)	0.51 (0.49-0.52)
Positive LR	3.27 (2.64-4.05)	2.48 (1.68-3.68)	2.37 (2.12-2.66)	1.49 (1.29-1.73)	2.38 (2.26-2.51)	1.85 (1.71-2.00)
Negative LR	0.69 (0.64-0.75)	0.74 (0.66-0.84)	0.10 (0.07-0.15)	0.13 (0.05-0.31)	0.17 (0.11-0.25)	0.18 (0.09-0.37)
DOR	4.7 (3.6-6.2)	3.3 (2.0-5.5)	22.9 (15.0-34.9)	11.7 (4.4-31.1)	14.3 (9.1-22.5)	10.2 (4.7-22.2)

Table E. Flowchart Shortness of breath in adults

	Erasmus MC		Maasstad		Fernando Fonseca	
	All adults n=1110	**Elderly n=382**	**All adults n=2312**	**Elderly n=1166**	**All adults n=10,907**	**Elderly n=6408**
High urgent patients, n (%)	237 (21.4)	101 (26.4)	955 (41.3)	604 (51.8)	4495 (41.2)	3252 (50.7)
Absolute classification (%)						
Correct triage	608 (54.8)	198 (51.8)	919 (39.7)	485 (41.6)	4225 (38.7)	2118 (33.1)
Overtriage	273 (24.6)	84 (22.0)	1260 (54.5)	605 (51.9)	5706 (52.3)	3796 (59.2)
Undertriage	229 (20.6)	100 (26.2)	133 (5.8)	76 (6.5)	976 (8.9)	494 (7.7)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.53 (0.46-0.60)	0.48 (0.39-0.58)	0.82 (0.75-0.87)	0.84 (0.76-0.90)	0.90 (0.82-0.94)	0.92 (0.84-0.96)
Specificity	0.86 (0.83-0.88)	0.81 (0.76-0.86)	0.62 (0.60-0.64)	0.52 (0.49-0.55)	0.59 (0.58-0.60)	0.50 (0.49-0.51)
Positive LR	3.68 (3.00-4.52)	2.59 (1.88-3.55)	2.15 (1.97-2.35)	1.74 (1.58-1.93)	2.20 (2.05-2.36)	1.83 (1.71-1.97)
Negative LR	0.55 (0.47-0.64)	0.64 (0.53-0.78)	0.29 (0.21-0.40)	0.31 (0.20-0.47)	0.18 (0.10-0.31)	0.16 (0.08-0.35)
DOR	6.7 (4.8-9.4)	4.1 (2.5-6.6)	7.3 (4.9-10.9)	5.7 (3.4-9.6)	12.6 (6.7-23.5)	11.4 (4.9-26.2)

Table F. Flowchart Headache

	Erasmus MC		Maasstad		Fernando Fonseca	
	All adults n=889	**Elderly n=163**	**All adults n=730**	**Elderly n=273**	**All adults n=7970**	**Elderly n=1874**
High urgent patients, n (%)	194 (21.8)	62 (38.0)	219 (30.0)	117 (42.9)	1228 (15.4)	375 (20.0)
Absolute classification (%)						
Correct triage	468 (52.6)	91 (55.8)	268 (36.7)	116 (42.5)	4017 (50.4)	859 (45.8)
Overtriage	300 (33.7)	39 (23.9)	436 (59.7)	145 (53.1)	2856 (35.8)	797 (42.5)
Undertriage	121 (13.6)	33 (20.2)	26 (3.6)	12 (4.4)	1097 (13.8)	218 (11.6)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.56 (0.49 to 0.63)	0.62 (0.50 to 0.72)	0.76 (0.55 to 0.89)	0.81 (0.57 to 0.93)	0.78 (0.55 to 0.91)	*
Specificity	0.88 (0.86 to 0.91)	0.81 (0.72 to 0.88)	0.71 (0.68 to 0.75)	0.60 (0.53 to 0.65)	0.85 (0.84 to 0.86)	
Positive LR	4.86 (3.82 to 6.18)	3.26 (2.05 to 5.19)	2.66 (2.04 to 3.47)	2.01 (1.52 to 2.65)	5.09 (3.96 to 6.56)	
Negative LR	0.50 (0.42 to 0.58)	0.47 (0.35 to 0.64)	0.33 (0.16 to 0.72)	0.31 (0.11 to 0.88)	0.26 (0.11 to 0.62)	
DOR	9.8 (6.8 to 14.1)	6.9 (3.4 to 14.0)	8.0 (2.9 to 22.1)	6.4 (1.8 to 22.9)	19.4 (6.4 to 59.1)	

**≤10 high urgent patients available for analysis*

S5 File. Diagnostic performance of the MTS for the most commonly used MTS discriminators, as determined by the 3-category reference standard

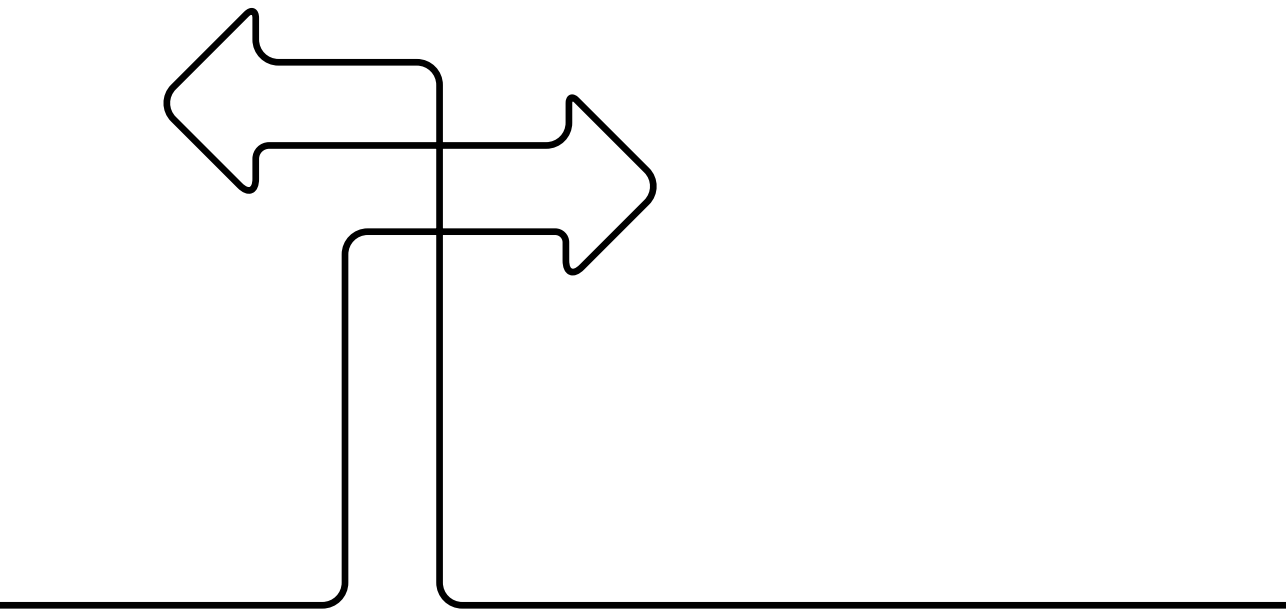
Table A. Consciousness discriminators

	Erasmus MC		Maasstad		Fernando Fonseca	
	All adults n=592	**Elderly n=135**	**All adults n=617**	**Elderly n=277**	**All adults n=4451**	**Elderly n=2553**
Absolute classification (%)						
Correct triage	260 (43.9)	65 (48.1)	188 (30.5)	104 (37.5)	1133 (25.5)	530 (20.8)
Overtriage	269 (45.4)	48 (35.6)	423 (68.6)	168 (60.6)	3315 (74.5)	2021 (79.2)
Undertriage	63 (10.6)	22(16.3)	6 (1.0)	5 (1.8)	3 (0.1)	2 (0.1)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.64 (0.57-0.71)	0.61 (0.48-0.73)	0.90 (0.80-0.95)	0.86 (0.71-0.94)	0.98 (0.96-0.99)	0.98 (0.94-1.00)
Specificity	0.78 (0.73-0.81)	0.60 (0.49-0.70)	0.59 (0.55-0.63)	0.55 (0.48-0.61)	0.43 (0.42-0.45)	0.32 (0.31-0.34)
Positive LR	2.87 (2.32-3.54)	1.54 (1.10-2.18)	2.21 (1.94-2.52)	1.89 (1.55-2.29)	1.73 (1.68-1.79)	1.46 (1.40-1.51)
Negative LR	0.46 (0.38-0.57)	0.64 (0.44-0.93)	0.17 (0.08-0.36)	0.26 (0.12-0.59)	0.04 (0.01-0.11)	0.05 (0.01-0.21)
DOR	6.2 (4.2-9.1)	2.4 (1.2-4.9)	13.1 (5.5-30.9)	7.2 (2.7-19.2)	49.0 (15.6-153.5)	27.4 (6.8-111.4)

Table B. Temperature discriminators

	Erasmus MC		Maasstad		Fernando Fonseca	
	All adults n=788	**Elderly n=217**	**All adults n=441**	**Elderly n=195**	**All adults n=2305**	**Elderly n=383**
Absolute classification (%)						
Correct triage	481 (61.0)	148 (68.2)	310 (70.3)	147(75.4)	1392 (60.4)	229 (59.8)
Overtriage	155 (19.7)	30 (13.8)	85 (19.3)	26 (13.3)	570 (24.7)	102 (26.6)
Undertriage	152 (19.3)	39 (18.0)	46 (10.4)	22 (11.3)	343 (14.9)	52 (13.6)
Diagnostic accuracy (95% confidence interval)						
Sensitivity	0.19 (0.13-0.27)	0.17 (0.07-0.34)	0.19 (0.11-0.31)	0.21 (0.10-0.40)	0.42 (0.19-0.68)	*
Specificity	0.96 (0.94-0.97)	0.96 (0.92-0.98)	0.94 (0.91-0.96)	0.91 (0.86-0.94)	0.98 (0.97-0.99)	
Positive LR	4.87 (2.91-8.15)	4.45 (1.51-13.12)	3.22 (1.66-6.25)	2.39 (1.01-5.63)	21.23 (10.24-44.03)	
Negative LR	0.84 (0.77-0.92)	0.87 (0.74-1.02)	0.86 (0.75-0.98)	0.86 (0.71-1.05)	0.60 (0.37-0.96)	
DOR	5.8 (3.2-10.4)	5.1 (1.5-17.4)	3.8 (1.7-8.2)	2.8 (1.0-7.9)	35.7 (10.9-116.7)	

*≤10 high urgent patients available for analysis



Chapter 4

The performance of triage systems in emergency care: a systematic review and meta-analysis

Joany M Zachariasse, Vera van der Hagen, Nienke Seiger, Kevin Mackway-Jones, Mirjam van Veen, Henriëtte A Moll

BMJ Open. 2019 May 28;9(5):e026471

ABSTRACT

Objective To assess and compare the performance of triage systems for identifying high and low-urgency patients in the emergency department (ED).

Design Systematic review and meta-analysis.

Data sources EMBASE, Medline OvidSP, Cochrane central, Web of science and CINAHL databases from 1980 to 2016 with the final update in December 2018.

Eligibility criteria Studies that evaluated an emergency medical triage system, assessed validity using any reference standard as proxy for true patient urgency and were written in English. Studies conducted in low(er) income countries, based on case scenarios or involving less than 100 patients were excluded.

Review methods Reviewers identified studies, extracted data and assessed the quality of the evidence independently and in duplicate. The Quality Assessment of studies of Diagnostic Accuracy included in Systematic Reviews -2 checklist was used to assess risk of bias. Raw data were extracted to create 2×2 tables and calculate sensitivity and specificity. ED patient volume and casemix severity of illness were investigated as determinants of triage systems' performance.

Results Sixty-six eligible studies evaluated 33 different triage systems. Comparisons were restricted to the three triage systems that had at least multiple evaluations using the same reference standard (Canadian Triage and Acuity Scale, Emergency Severity Index and Manchester Triage System). Overall, validity of each triage system to identify high and low-urgency patients was moderate to good, but performance was highly variable. In a subgroup analysis, no clear association was found between ED patient volume or casemix severity of illness and triage systems' performance.

Conclusions Established triage systems show a reasonable validity for the triage of patients at the ED, but performance varies considerably. Important research questions that remain are what determinants influence triage systems' performance and how the performance of existing triage systems can be improved.

INTRODUCTION

Overcrowding of emergency departments (EDs) is a universal and ever-increasing problem.[70-72] Therefore, most EDs have a triage system in place to facilitate the prioritisation of patients. In recent years, several formal triage scales have been developed to standardise the approach to triage. These include amongst others the Australasian Triage Scale, the Canadian Triage and Acuity Scale (CTAS), the Emergency Severity Index (ESI) and the Manchester Triage System (MTS).[4, 10-12]

It is important to evaluate the performance of triage systems for their ability to accurately distinguish between both high and low-urgency patients. The correct classification of high-urgency patients is related to patient safety, because misclassification of high-urgency patients to a low-urgency level causes delay in diagnosis and treatment, potentially leading to morbidity or mortality. The correct classification of low-urgency patients increases efficiency of the ED flow and reduces waiting times for the truly high-urgency ED visits.

Research regarding the performance of triage systems mainly consists of observational studies in heterogeneous populations using a variety of reference standards. Previous reviews have primarily described the results of these individual studies without combining and interpreting the evidence into overall conclusions.[73-75] A systematic appraisal of the performance of commonly used triage systems can inform clinicians and policy-makers about the safety and efficacy of available triage systems and provide insights on which triage system is safe and efficient to use. Moreover it can highlight gaps in current research and propose directions for future studies.

The aim of this systematic review is to provide a comprehensive overview of current evidence on the performance of triage systems. We assessed and compared the performance of the most commonly used triage systems for the prioritisation of high and low-urgency patients at the ED, compared with any reference standard that is a proxy of true patient urgency. Furthermore, we aimed to investigate whether patient volume and casemix at the ED are determinants of triage systems' performance.

METHODS

Search strategy

Meta-analysis of Observational Studies in Epidemiology guidelines were followed for the conduct of this study.[76] We conducted a systematic review using a broad search strategy to identify all studies assessing the performance of triage systems in emergency care when compared with any reference standard that is a proxy of true patient urgency. A search strategy was developed by a health sciences librarian and included medical subject headings and text words related to triage, emergency care and validity (Appendix

1). We searched EMBASE, Medline OvidSP, Cochrane central, Web of science and CINAHL databases from 1980 to 2016 with the final update in December 2018.

Study eligibility

Studies were selected that assess the performance of triage systems in emergency care with a defined outcome measure as a proxy of true patient-urgency. We selected studies based on the following PICO:

Population: We included studies evaluating triage in the unselected group of patients attending the ED. We excluded studies restricted to specific patient subgroups (such as patients with specific diseases).

Interventions: We included any studies assessing ED triage systems, defined as any tool aimed to classify patients at the ED based on the urgency or severity of their condition. We did not include studies evaluating trauma triage systems or early warning scores.

Comparators: Since no golden standard for the evaluation of triage systems exists, we included all studies evaluating the performance of triage systems using one or more defined reference standard as a proxy for true patient urgency.

Outcome: We defined outcome as the sensitivity and specificity of the triage system for the identification of high-urgent and low-urgent patients. A-priori, we selected mortality at the ED and Intensive Care Unit (ICU) admission after the ED visit as reference standard for high-urgency and discharge home after the ED visit as the reference standard for low urgency. We additionally considered any other reference standard with sufficient evaluations.

Letters, abstracts, reviews, conference proceedings, and case reports were excluded as well as studies not written in English. We excluded studies with less than 100 patients and studies based on case scenarios because these studies have of a high risk of bias. Moreover, we excluded studies conducted in low or lower-income economies.[77] The unique characteristics of EDs in these countries, including the number of patients, epidemiology of diseases and available resources, make study results difficult to compare to middle or higher income countries. Two reviewers (MvV and NS or JZ and VvdH) independently assessed eligibility for inclusion. Disagreements in article selection were resolved through discussion.

Data extraction

Two reviewers (JZ and VvdH) independently extracted data from each of the included studies. A single study could consist of multiple "triage evaluations", defined as the analysis of a single triage system in a single age group. Predefined age groups were 1) children; 2) adults or a combination of age groups; and 3) elderly. For each of the triage evaluations, the reviewers extracted the total number of included patients, the number of patients in each of the urgency categories of the triage system, the type of reference standard used, and the number of patients with a positive reference standard in each urgency category. If studies were based on overlapping data, we used the results from

the most recent publication. For descriptive purposes we also collected data on study design and methods, patient demographics, and characteristics of the settings in which the study was performed.

Quality assessment

Two reviewers independently assessed quality of the selected articles using the Quality Assessment of studies of Diagnostic Accuracy included in Systematic Reviews (QUADAS-2) checklist.[78] The QUADAS-2 evaluates four domains: patient selection, index test, reference standard, and flow and timing. Each domain is assessed in terms of risk of bias, and the first three domains also in terms of applicability. Because triage systems have some specific features as compared with other diagnostic tests, we adjusted the "reference standard domain" to make it applicable to our research question. We did not appraise whether the reference standard was interpreted without knowledge of the result of the index test, because this is unlikely when triage is applied in routine care. We did evaluate, however, whether data on the outcome was collected blinded to the result of the index test. Moreover, we did not judge the applicability of the reference standard because there is no consensus on this topic.[79] Therefore, we included all studies with reference standard that were a proxy for patient urgency. Any disagreements between reviewers were resolved in a consensus meeting.

Data analysis

We used descriptive analyses to provide an overview of the available evidence on triage systems. Further analyses were restricted to triage systems that underwent at least three evaluations with the same reference standard. The primary outcome of our review was sensitivity and specificity of each of the triage systems for the identification of high and low-urgency patients. Because there is no golden standard to determine "true" patient urgency, we a priori selected three reference standards as proxy for patient urgency. We considered mortality at the ED and ICU admission after the ED visits as reference standard for high urgency, and discharge after the ED visit (ie, patients not admitted to hospital) as reference standard for low urgency. Although these measures are not perfect, they approximate the desired outcome: most patients who die at the ED or require ICU admission are of high urgency, while most patients who are discharged after the ED visit are not. Moreover, these measures are suitable for the analyses of large datasets and commonly reported in research on triage systems.[79] In addition, we considered any other reference standard with sufficient evaluations in the same triage system.

We calculated two by two tables of triage system against the reference standard for each individual study. Because triage systems are ordinal scales, we dichotomized the urgency categories into a high-urgency and low-urgency group. High urgency was defined as triage urgency level 1 (three-level systems) or triage urgency level 1 and 2 (four-level and five-level systems).

We calculated sensitivities and specificities, and presented the results as forest plots. We aimed to summarize the diagnostic accuracy data using a bivariate random effects model, but due to the substantial heterogeneity between studies this was not possible. For clinical practice and for benchmarking purposes, we calculated the proportion of patients with a positive reference standard per urgency category. These results are displayed in a bar chart to enable comparison between studies and between triage systems.

We hypothesized that ED patient volume and casemix severity of illness were determinants of triage systems' performance. Therefore we decided that if a sufficient number of studies were identified, we would investigate the effect of these determinants on triage systems' performance using subgroup analyses. We considered annual ED census as a marker of patient volume and the percentage of hospitalized patients as a marker of casemix severity of illness.

Computations were carried out with SPSS Statistics V.21.0 and figures were created using Review Manager V.5.3 or R V.3.2.0.[80-82]

Patient and Public Involvement

Patients and public were not involved in this study.

RESULTS

A total of 12,684 papers were identified in the electronic search, of which 66 were included in the final selection (figure 1).

The majority of studies were conducted in tertiary or university hospitals (n=46; 70%) and conducted in Europe/Central Asia or North America (n=45; 68%). Forty-nine (74%) were single-centre studies. A complete overview of the selected studies is presented in appendix 2.

Forty-four studies (67%) had a high risk of bias in at least one domain, and 17 studies (26%) had a high risk of bias in two or more domains (figure 2 and appendix 3). The most common causes of concern were application of multiple triage systems for the same patient by the same nurse or by multiple nurses without blinding, retrospective retrieval of reference standard information without blinding for the triage outcome, and substantial amounts of missing data. In 11 studies (17%) there were concerns regarding applicability.

Figure 1. Flow diagram of study selection process

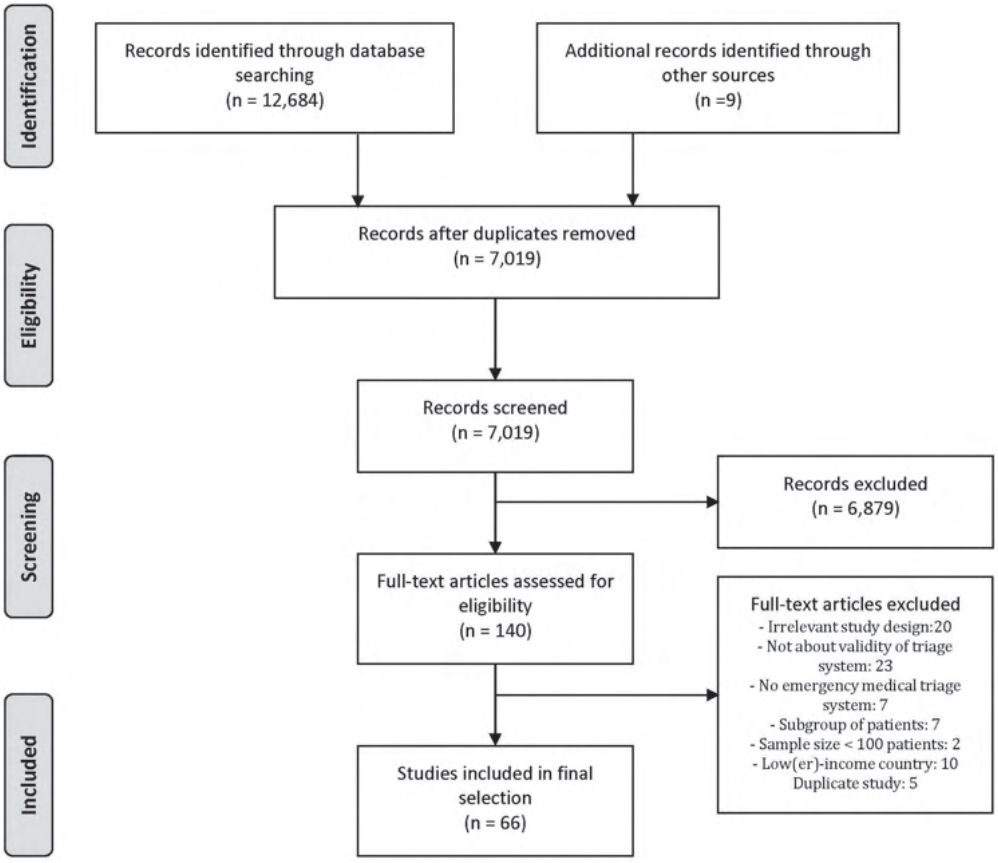
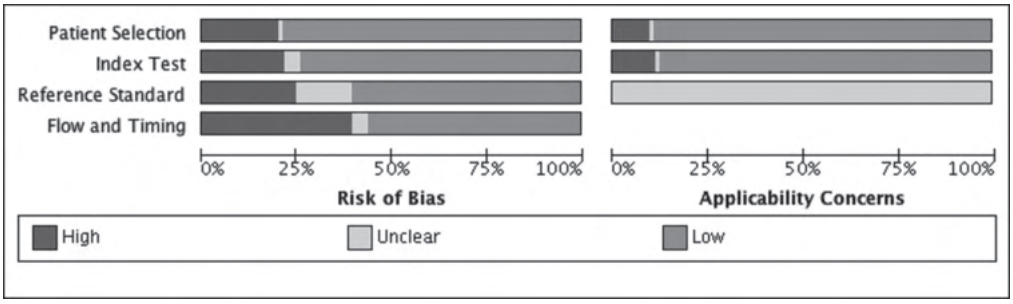


Figure 2. Risk of bias of included studies



Triage systems

A total of 33 different triage systems were evaluated. The most commonly evaluated triage systems were the ESI (n=22), the MTS (n=15), and the CTAS (n=13). Other triage

systems included the Taiwan Triage System (n=4), Australasian Triage Scale (n=3), South African Triage Scale (n=3), Netherlands Triage System (n=2), and Soterion Rapid Triage System (n=2). For 25 triage systems only one evaluation was published. These included nine local or informally structured triage systems. The median sample size was 1,496 in children (range: 510 to 550,940), 1,447 in adults (range: 100 to 316,622), and 929 in elderly (range: 773 to 1,903). In total, 89 individual triage evaluations were reported: 34 (38%) in children, 52 (58%) in adults, a combination of age groups or an unspecified population, and 3 (3%) in elderly.

Reference standards

A variety of reference standards were used and the majority of studies reported multiple reference standards (Appendix 4). Twelve studies used mortality at the ED as a reference standard, 13 studies ICU admission and 47 studies hospital admission. Other commonly reported reference standards were length of stay at the ED (27 studies), resource use at the ED according to the ESI criteria (14 studies), expert opinion (9 studies) and costs (8 studies). Because definitions of time to mortality, resource use and expert opinion were not consistent across studies, these results could not be compared. Moreover, length of stay at the ED and costs are outcome measures that are strongly dependent on ED characteristics, and could therefore not be used to make a comparison between studies.

The most commonly evaluated triage systems

We will further restrict our analyses to the triage systems with at least three evaluations using the same reference standard. This final selection includes studies evaluating the ESI, CTAS and MTS. Characteristics of these triage systems and a summary of the available evidence are presented in table 1. The ESI, CTAS and MTS were all evaluated in settings with a different patient volume as indicated by annual hospital census, and a different case-mix as indicated by percentage of hospitalisation. For each of these triage systems, the majority of studies had risk of bias in at least one domain.

Accuracy of triage systems to identify high urgent patients

Mortality at the ED was reported in seven evaluations of our final sample: five evaluations in adults, and two in children. Because of this low number of studies and the very low reported mortality rates (on average 0.2% in adults and <0.01% in children) it was not possible to perform comparative analyses.

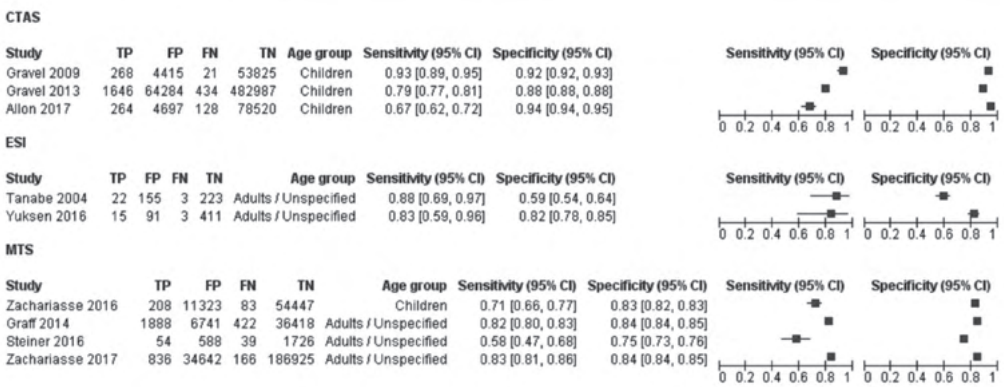
ICU admission was reported in five evaluations in adults (two ESI, three MTS) and four in children (three CTAS, one MTS). Overall, sensitivity for ICU admission was moderate to good, ranging from 0.58 (95%CI 0.48 to 0.68) to 0.88 (95%CI 0.70 to 0.96) in adults and 0.71 (95%CI 0.66 to 0.77) to 0.93 (95%CI 0.89 to 0.95) in children. A clear difference in performance between the triage systems was not visible (figure.3).

Table 1. Evidence summary of the most commonly used triage systems

	Canadian Triage And Acuity Scale (CTAS)	Emergency Severity Index (ESI)	Manchester Triage System (MTS)
Triage system characteristics			
Description	List, based on presenting signs and symptoms	Flowchart, based on physical signs and expected resource use	Multiple flowcharts, based on presentational signs and symptoms
Number of levels	5	5	5
Classification and waiting time	Level I, Immediate Level II, 15 min Level III, 30 min Level IV, 60 min Level V, 2 hrs	Immediate Emergent, 14 min Urgent, 60 min Semi-urgent, 2 hrs Non-urgent, 24 hrs	Immediate Very urgent, 10 min Urgent, 60 min Standard, 2 hrs Non-urgent, 4 hrs
Quantity of evidence			
Total number of evaluations	13	21	15
Evaluations in children	9	4	7
Evaluations in elderly	1	2	0
Diversity of evidence (range)			
Number of hospitals per study	1-12	1-7	1-4
Inclusions per study	481-550,940	180-37,974	872-31,622
Hospital census	10,000-75,000	10,000-90,000	7,000-190,000
Hospitalisation rate	8%-47%	10%-62%	5-33%
Risk of bias			
High risk of bias in at least one domain	54%	81%	67%
High risk of bias in >1 domain	15%	23%	13%

Regardless of the triage system used, most of the ICU admitted patients were allocated to one of the two highest triage categories (Appendix 5). The exact proportion of ICU admitted patients in each triage category was highly variable, even within studies evaluating the same triage system. For example, the proportion of ICU admitted adults in MTS category 1 ranged from 21% to 79%. The number of studies, however, was too small to assess whether this variability was present in all triage systems and whether this could be explained by study or setting related factors.

Figure 3. Sensitivity and specificity of triage systems for identifying high urgency patients as defined by ICU admission



CTAS = Canadian Triage and Acuity Scale; TP = true positive; FP = false positive; FN = false negative; TN = true negative; ESI = Emergency Severity Index; MTS = Manchester Triage System

Accuracy to identify low urgent patients

Hospital admission or discharge after the ED visit was reported as a reference standard in 14 evaluations in adults and 15 in children. Overall, specificity of the triage systems to accurately classify patients discharged home as low urgent ranged from 0.64 (95%CI 0.62 to 0.66) to 0.98 (95%CI 0.95 to 0.99) in adults and 0.69 (95%CI 0.66 to 0.72) to 0.96 (95%CI 0.94 to 0.98) in children (Figure 4). Again, sensitivities and specificities were highly variable within each of the triage systems. None of the triage systems showed a marked better specificity compared with the others.

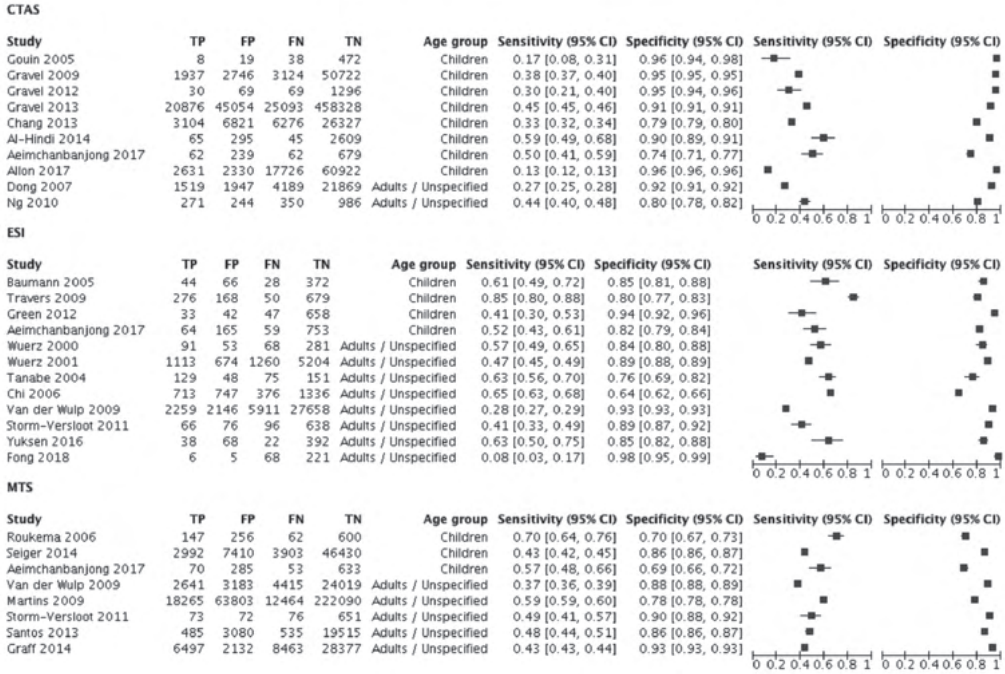
The proportion of patients discharged after the ED visit increased from the higher to the lower urgency categories in all triage systems (Appendix 5). Again, there was a large variability within triage systems and substantial overlap between triage systems. In adults, the MTS seemed to have a higher variability compared with the other triage systems, but in children variability was greater for the CTAS.

The only additional reference standard with sufficient evaluations was resource use according to the ESI criteria (Appendix 6).

Direct comparison of triage systems

A total of 13 studies directly compared two or more triage systems. Most of these studies, however, were assessed as having a high risk of bias in the index test domain, because triage was performed by the same nurse or without blinding. Performing triage, while using different triage systems sequentially, is likely to reduce the differences between triage systems. Therefore, these results should be interpreted with caution.

Figure 4. Sensitivity and specificity of triage systems for identifying low urgency patients as defined by discharge home after the ED visit



CTAS = Canadian Triage and Acuity Scale; TP = true positive; FP = false positive; FN = false negative; TN = true negative; ESI = Emergency Severity Index; MTS = Manchester Triage System

In 10 studies, an established five-level triage system was compared with a local or informally structured triage system with three or four levels. Seven of these studies reported that the five-level triage system provided better discrimination or better sensitivities and specificities than the local triage system and should be preferred.[83-89] One study in children found that the local triage system performed better than the established triage system (CTAS).[90]

Two studies comparing the ESI with the MTS in adults found that sensitivities and specificities for hospital admission and the prediction of mortality were largely similar.[91, 92] In one of these studies, the MTS undertriaged a smaller proportion of patients compared to the ESI (8.3% versus 13.5%) at the cost of a larger proportion of "overtriage".[91] One study in adults observed no statistically significant difference between the CTAS and ESI regarding the prediction of emergency department resource utilization and immediate patient outcomes.[93] One study in children compared the ATS and ESI and found similar sensitivities and specificities for the identification of patients requiring hospital admission.[94] One single-centre study in children compared five triage systems (ATS, CTAS, ESI, MTS and a local triage system called the Ramathibodi Triage System) and concluded that the ESI showed the best validity for predicting hospital admission (Area Under the

Curve 0.78, 95%CI 0.74 to 0.81).[95] In this study, the local triage system showed the highest sensitivity (50%) and the ATS the highest specificity (94%).

Determinants of triage systems' performance

The number of studies per triage system was too small to perform subgroup analyses based on annual census or percentage hospitalisation. As an explorative analysis, we ordered all selected studies that used hospital admission as a reference standard based on annual census, and percentage hospitalisation (Appendix 7). There was no clear association between patient volume or casemix and triage systems' sensitivity and specificity. A lower specificity for hospitals with the largest annual census and highest percentage hospitalisation could not be ruled out, but requires a larger number of studies.

DISCUSSION

In a systematic review of 66 observational studies evaluating triage systems, we found that numerous different triage systems are being used but that many lack a rigorous evaluation. The most commonly used and evaluated triage systems, CTAS, ESI and MTS, show a moderate to good validity to identify high and low-urgency patients. Their performance, however, is highly variable and differences in study design, study populations and reference standards make a comparison of the available evidence difficult. Although based on a limited number of studies, no clear association between patient volume and casemix severity of illness could be found.

Strengths and weaknesses

This is the first study that evaluates the performance of triage systems in a meta-analysis. Previous reviews have merely described the results of the individual studies without synthesising the evidence. Moreover, none of the published reviews looked at other factors that determine triage systems' performance, such as ED characteristics to compare evidence from different studies.[73-75, 96-99]

Our review is based on a comprehensive search developed with a research librarian, includes duplicate assessment of eligibility and risk of bias, and duplicate data abstraction. Furthermore, the research question is based on a relevant and practical clinical issue. Triage systems are used worldwide to prioritise patients in the ED, but robust evidence on their performance is lacking.

The results of this review, however, should be interpreted taking into consideration the limitations of the underlying evidence. We included 66 studies in our review, but the majority of the 33 triage systems were evaluated by only one study. Therefore, we

could only evaluate three most frequent used triage systems: CTAS, ESI and MTS. Even for these commonly used triage systems, few evaluations were available due to the variety of reference standards. Although the triage evaluations included the whole age spectrum, studies targeted at elderly patients were scarce. It is important to evaluate triage systems' performance separately for the most vulnerable populations at the ED, specifically children and elderly. In these patient groups, the spectrum of disease, the presence of non-specific signs and symptoms and progression of disease course differs from that in adult patients.

We are not aware of any (randomized) controlled trial that investigates the effect of triage on patient outcome. Therefore we conducted a systematic review of observational studies. Comparing observational studies is challenging because the effect of a triage system cannot be assessed independently of its context. Likely, other factors such as ED and hospital characteristics or local practices and training have influenced the results of the included studies. We chose to display the results of different studies in a forest plot. Because of the limited number of studies and heterogeneity of the study populations, it is difficult to compare the results from different triage systems and these plots should be interpreted with caution.

We aimed to explore heterogeneity between studies and more specifically the effect of differences in patient load and casemix severity of illness. Unfortunately, due to the small number of studies per triage systems, we could not draw strong conclusions about the relation between these factors and triage systems' performance. There are more potential factors that could affect triage systems' validity, such as the local infrastructure, the experience and training of the triage nurse, the presence of a computerized triage application or variations in disease epidemiology. Moreover, since most included studies had risk of bias in at least one domain, we cannot rule out that study design and methodological quality have led to heterogeneity of the results as well.

We predefined mortality at the ED and ICU admission as reference standards for high patient urgency and discharge home as a reference standard for low patient urgency. Due to the relatively low number of studies reporting mortality and the low mortality rate of patients in the ED, we could not use it as a reference standard. ICU and hospital admission are feasible reference standards for large study populations, and theoretically, criteria for ICU and hospital admission should be reasonably comparable between settings. It is possible, however, that ED and hospital characteristics and local practices result in differences in the decision to admit a patient between EDs.

We restricted our review to triage systems in high-income and higher-middle income countries. We applied this selection because EDs in lower income countries have their own unique characteristics and challenges. Several recently published studies have addressed triage in low-income settings.[100, 101]

Implications and future research

Our review identified 33 triage systems for which at least one evaluation was published. Probably, there are more triage systems in use, which are not formally evaluated. There are several advantages, not addressed in this review, of using an established triage system over a local triage scale. Beside that performance of the most commonly used triage systems is known, they have a formal governance structure and undergo regular updates. Moreover, there are standardised implementation guidelines and training programs available.[20, 102, 103] The CTAS, ESI and MTS all show a reasonable performance for triage at the ED. Our results do not suggest that one of the established triage systems should be preferred over the other.

Our review indicates that large variation of performance exists even in studies assessing the same triage system. This suggests that other factors influence triage systems' performance. Consequently, generalisability of individual studies evaluating a triage system is low and a study on triage validity in one setting may not apply to a setting with different characteristics or in a different healthcare system. Our review demonstrates that the majority of studies evaluating triage systems were conducted in a single centre. Furthermore, most multicentre studies provided only pooled results. Yet, multicentre studies using similar study designs and reference standard definitions are needed to evaluate the range of triage systems' performance in different settings. More importantly, these studies can provide valuable insights in determinants of triage systems' performance and areas of improvement.

The extensive use of triage systems in clinical practice contrasts with the limited number of studies evaluating their performance. Triage systems were typically developed based on expert opinion and implemented out of clinical necessity. They are mostly used in their country or region of origin: the MTS is widely used in the UK and Europe, the ESI in the USA, and the CTAS in Canada and in French-speaking countries.[17] Since most EDs already have experience with a certain triage system and some triage systems are recommended by national guidelines it could be worthwhile shifting away from the focus on triage systems' performance towards the improvement of the established triage systems. Our review suggests that there is room for improvement of all triage systems regarding both the correct identification of high-urgency and low-urgency patients.

The most commonly used triage systems, CTAS, ESI and MTS, have a reasonable validity for the triage of patients at the ED. Important research questions that remain are what determinants influence a triage systems' performance and how the performance of existing triage systems can be improved.

ACKNOWLEDGEMENTS

We thank Wichor Bramer, biomedical information specialist in Erasmus University Medical Centre, Rotterdam, for his help with developing the search strategy

SUPPORTING INFORMATION

Available as online web appendix on the website of BMJ Open:
- Appendix 2. Selected studies
- Appendix 3. Risk of bias assessment

Appendix 1. Search strategy

Embase

(triage* OR ((acuity OR severit* OR priorit* OR critical*) NEAR/3 (scale* OR level* OR index* OR score* OR measure* OR assessment*))):ab,ti AND ('emergency ward'/de OR 'emergency medicine'/de OR 'emergency care'/exp OR (((emergen* OR acute) NEAR/3 (department* OR ward* OR unit* OR room* OR care))):ab,ti) AND (reliability/exp OR reproducibility/de OR 'validation process'/de OR 'validation study'/de OR accuracy/de OR 'sensitivity and specificity'/exp OR 'diagnostic accuracy'/de OR evaluation/de OR validity/exp OR (reliab* OR reproducib* OR validation OR validaty OR consisten* OR variabilit* OR accura* OR 'intra observer' OR intraobserver OR sensitivity OR specificity):ab,ti OR evaluat*:ti) AND [english]/lim

Medline Ovid

(Triage/ OR (triage* OR ((acuity OR severit* OR priorit* OR critical*) ADJ3 (scale* OR level* OR index* OR score* OR measure* OR assessment*))).ab,ti.) AND (exp "Emergency Service, Hospital"/ OR "emergency medicine"/ OR "emergency care"/ OR (((emergen* OR acute) ADJ3 (department* OR ward* OR unit* OR room* OR care))).ab,ti.) AND (Reproducibility of Results/ OR Validation Studies.pt. OR exp "sensitivity and specificity"/ OR Evaluation Studies.pt. OR (reliab* OR reproducib* OR validation OR validaty OR consisten* OR variabilit* OR accura* OR "intra observer" OR intraobserver OR sensitivity OR specificity). ab,ti. OR evaluat*.ti.) AND english.la.

Cochrane central

(triage* OR ((acuity OR severit* OR priorit* OR critical*) NEAR/3 (scale* OR level* OR index* OR score* OR measure* OR assessment*))):ab,ti AND ((((emergen* OR acute) NEAR/3 (department* OR ward* OR unit* OR room* OR care))):ab,ti) AND ((reliab* OR reproducib* OR validation OR validaty OR consisten* OR variabilit* OR accura* OR 'intra observer' OR intraobserver OR sensitivity OR specificity):ab,ti OR evaluat*:ti)

Web-of-science

TS=(triage* OR ((acuity OR severit* OR priorit* OR critical*) NEAR/3 (scale* OR level* OR index* OR score* OR measure* OR assessment*))) AND TS=((((emergen* OR acute) NEAR/3

(department* OR ward* OR unit* OR room* OR care)))) AND (TS=(reliab* OR reproducib* OR validation OR validaty OR consisten* OR variabilit* OR accura* OR "intra observer" OR intraobserver OR sensitivity OR specificity) OR TI=(evaluat*)) AND LA=(English)

CINAHL

(MH Triage+ OR (triage* OR ((acuity OR severit* OR priorit* OR critical*) N3 (scale* OR level* OR index* OR score* OR measure* OR assessment*)))) AND (MH "Emergency Service"+ OR MH "emergency medicine"+ OR (((emergen* OR acute) N3 (department* OR ward* OR unit* OR room* OR care)))) AND (MH Reproducibility of Results+ OR MH Validity+ OR MH "sensitivity and specificity"+ OR MH Evaluation+ OR (reliab* OR reproducib* OR validation OR validaty OR consisten* OR variabilit* OR accura* OR "intra observer" OR intraobserver OR sensitivity OR specificity) OR TI (evaluat*))

Google Scholar

triage emergency|emergencies reliability|reproducibility|validation|validity| consistency| accuracy|sensitivity|specificity|interobserver|intraobserver

Appendix 4. Reference standards used in studies evaluating triage systems in the emergency department

Patient disposition and follow-up	Hospital admission	47 studies
	ICU admission	13 studies
	Follow-up at outpatient clinic or GP	1 study
Waiting times	Length of stay at the ED	27 studies
	Length of stay in hospital	5 studies
	Waiting time to physician	4 studies
	Other: Waiting time to nurse, Waiting time to examination, Waiting time to first lab results, Waiting time from arrival to treatment room, Waiting time from treatment room to discharge from ED, Waiting time for care	1 study each
Resource use	Resource use based on ESI criteria	14 studies
	Costs	8 studies
	Current Procedural Terminology (CPT) codes	2 studies
	Other	15 studies
Mortality	Mortality at the ED	12 studies
	In-hospital mortality	5 studies
	1 year mortality	2 studies
	Other: 30-day mortality, 60-day mortality, 6 month mortality	1 study each
Composite outcomes	"Immediate lifesaving interventions"	4 studies
	"MTS reference standard": Combination of abnormal physiologic parameters, life-threatening conditions, resource use, follow-up	4 studies
	"Critical or time-sensitive outcomes": in-hospital mortality, intensive care unit admission or transfer to operating room or catheterization suite	1 study
	"Actual clinical outcomes": Combination of life threatening condition, transfer to higher hospital, ICU, or hospital admission	1 study
	Composite outcome based on physiologic parameters, patient disposition and resource use	1 study
Physiologic parameters	Worthing Phsyiological Scoring System	1 study
	Pain score	1 study
Other	Expert opinion	9 studies
	Left without being seen	5 studies
	Patient or parent satisfaction	2 studies
	Level of prehospital care	1 study
	Sentinel diagnosis	1 study
	Pediatric risk of admission (PRISA) score	1 study

Appendix 5. Reference standard per urgency category

Figure 5.1 Proportion of ICU admissions per urgency category in the most commonly evaluated triage systems.

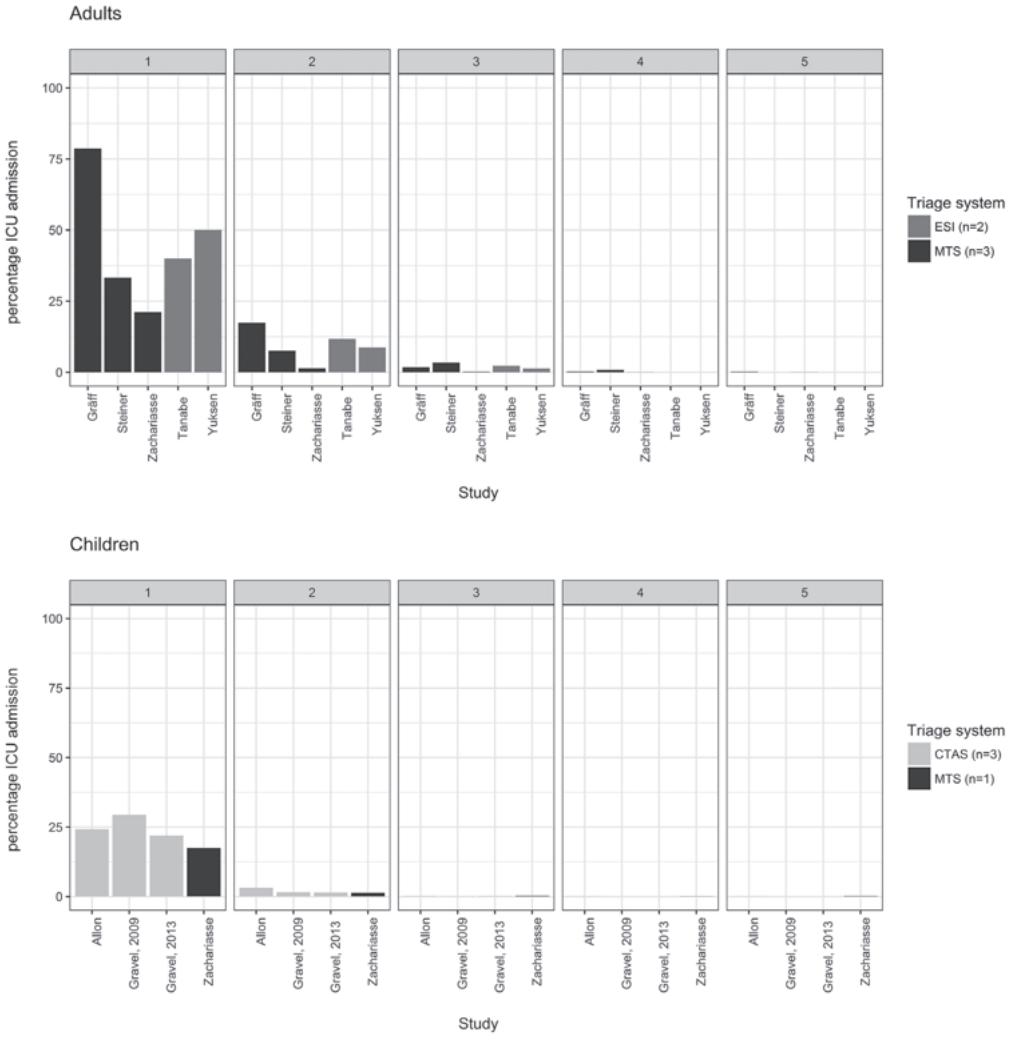
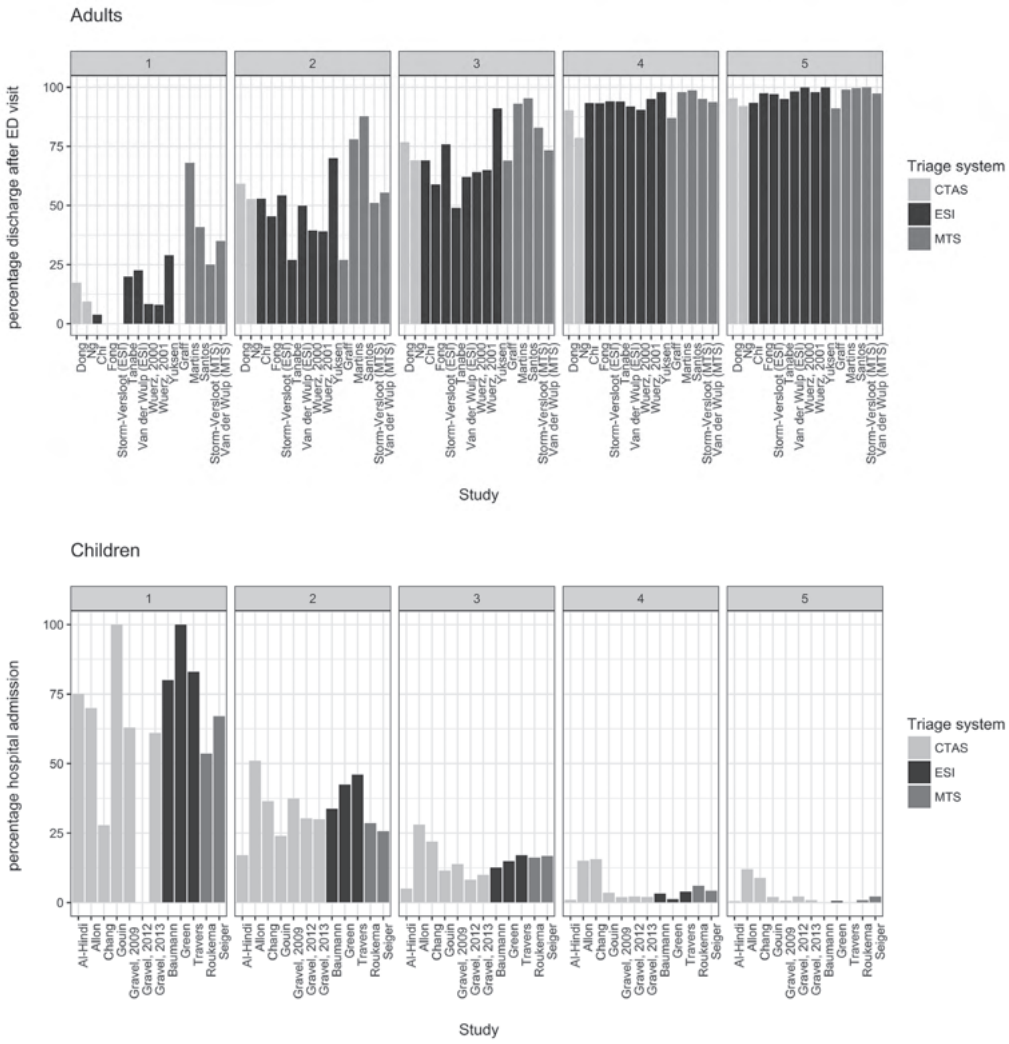


Figure 5.2 Proportion of patients discharged home per urgency category in the most commonly evaluated triage systems



Appendix 6. Other reference standards

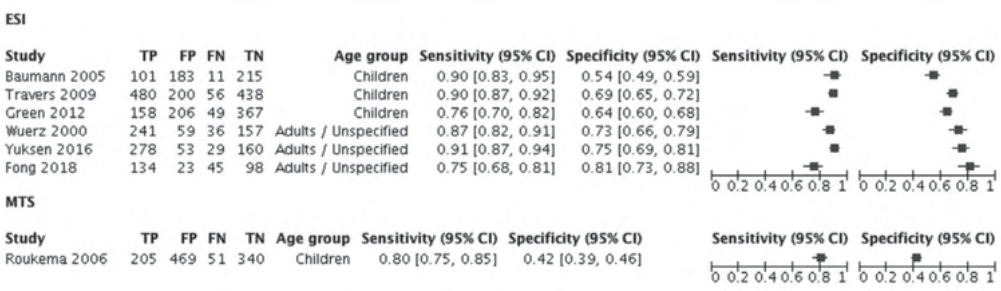
Other reference standards

14 studies used resource use according to ESI criteria as a reference standard of which 7 provided data to construct a 2x2 table and were conducted in one of the commonly used triage systems. One study evaluated resource use in the MTS, and all others studied the ESI.

Because the ESI states that resource prediction should only be used for the less acute patients, we used a different cut-off for dichotomization to calculate sensitivity and specificity. We dichotomized the triage categories into a high urgent group consisting of triage urgency levels 1, 2 and 3 with a low urgent group consisting of triage categories 4 and 5.

Overall, sensitivity of the triage systems to accurately classify patients with 2 or more resources in the highest urgency categories ranged from 0.75 (95%CI 0.68 to 0.81) to 0.91 (95%CI 0.87 to 0.94) in adults and from 0.76 (95%CI 0.70 to 0.82) to 0.90 (95%CI 0.87 to 0.92) in children (Fig. 7). Specificity of the triage system to accurately classify patients with 1 or 0 resources in the lowest two urgency categories ranged from 0.73 (95%CI 0.66 to 0.79) to 0.81 (95%CI 0.73 to 0.88) in adults and from 0.42 (95%CI 0.39 to 0.46) to 0.75 (95%CI 0.69 to 0.81) in children.

Fig 6.1. Forest plot evaluating sensitivity and specificity of triage systems using resource use according to ESI criteria as reference standard



ESI = Emergency Severity Index; TP = true positive; FP = false positive; FN = false negative; TN = true negative; MTS = Manchester Triage System

Appendix 7. Determinants of triage systems' performance

Figure 7.1. Forest plot evaluating sensitivity and specificity of triage systems for identifying low urgency patients as defined by discharge home after the ED visit, ordered by annual census as a marker of patient volume

Adults

Study	TP	FP	FN	TN	Annual Census	Sensitivity (95% CI)	Specificity (95% CI)
Graff 2014	6497	2132	8463	28377	23000.0	0.43 [0.43, 0.44]	0.93 [0.93, 0.93]
Storm-Versloot 2011	66	76	96	638	31000.0	0.41 [0.33, 0.49]	0.89 [0.87, 0.92]
Storm-Versloot 2011	73	72	76	651	31000.0	0.49 [0.41, 0.57]	0.90 [0.88, 0.92]
Yuksen 2016	38	68	22	392	33936.0	0.63 [0.50, 0.75]	0.85 [0.82, 0.88]
Van der Wulp 2009	2259	2146	5911	27658	38500.0	0.28 [0.27, 0.29]	0.93 [0.93, 0.93]
Van der Wulp 2009	2641	3183	4415	24019	38500.0	0.37 [0.36, 0.39]	0.88 [0.88, 0.89]
Wuerz 2001	1113	674	1260	5204	55000.0	0.47 [0.45, 0.49]	0.89 [0.88, 0.89]
Chi 2006	713	747	376	1336	58000.0	0.65 [0.63, 0.68]	0.64 [0.62, 0.66]
Wuerz 2000	91	53	68	281	60000.0	0.57 [0.49, 0.65]	0.84 [0.80, 0.88]
Dong 2007	1519	1947	4189	21869	67000.0	0.27 [0.25, 0.28]	0.92 [0.91, 0.92]
Tanabe 2004	129	48	75	151	70000.0	0.63 [0.56, 0.70]	0.76 [0.69, 0.82]
Santos 2013	485	3080	535	19515	100000.0	0.48 [0.44, 0.51]	0.86 [0.86, 0.87]
Martins 2009	18265	63803	12464	222090	128000.0	0.59 [0.59, 0.60]	0.78 [0.78, 0.78]

Children

Study	TP	FP	FN	TN	Annual Census	Sensitivity (95% CI)	Specificity (95% CI)
Aeimchanbanjong 2017	62	239	62	679	10000.0	0.50 [0.41, 0.59]	0.74 [0.71, 0.77]
Aeimchanbanjong 2017	64	165	59	753	10000.0	0.52 [0.43, 0.61]	0.82 [0.79, 0.84]
Aeimchanbanjong 2017	70	285	53	633	10000.0	0.57 [0.48, 0.66]	0.69 [0.66, 0.72]
Allon 2017	2631	2330	17726	60922	16722.0	0.13 [0.12, 0.13]	0.96 [0.96, 0.96]
Roukema 2006	147	256	62	600	30000.0	0.70 [0.64, 0.76]	0.70 [0.67, 0.73]
Chang 2013	3104	6821	6276	26327	42839.0	0.33 [0.32, 0.34]	0.79 [0.79, 0.80]
Green 2012	33	42	47	658	45000.0	0.41 [0.30, 0.53]	0.94 [0.92, 0.96]
Baumann 2005	44	66	28	372	53000.0	0.61 [0.49, 0.72]	0.85 [0.81, 0.88]
Gravel 2009	1937	2746	3124	50722	60000.0	0.38 [0.37, 0.40]	0.95 [0.95, 0.95]
Gouin 2005	8	19	38	472	65000.0	0.17 [0.08, 0.31]	0.96 [0.94, 0.98]
Al-Hindi 2014	65	295	45	2609	155000.0	0.59 [0.49, 0.68]	0.90 [0.89, 0.91]

TP = true positive; FP = false positive; FN = false negative; TN = true negative

Figure 7.2. Forest plot evaluating sensitivity and specificity of triage systems for identifying low urgency patients as defined by discharge home after the ED visit, ordered by percentage hospitalization as a marker of case-mix severity of illness

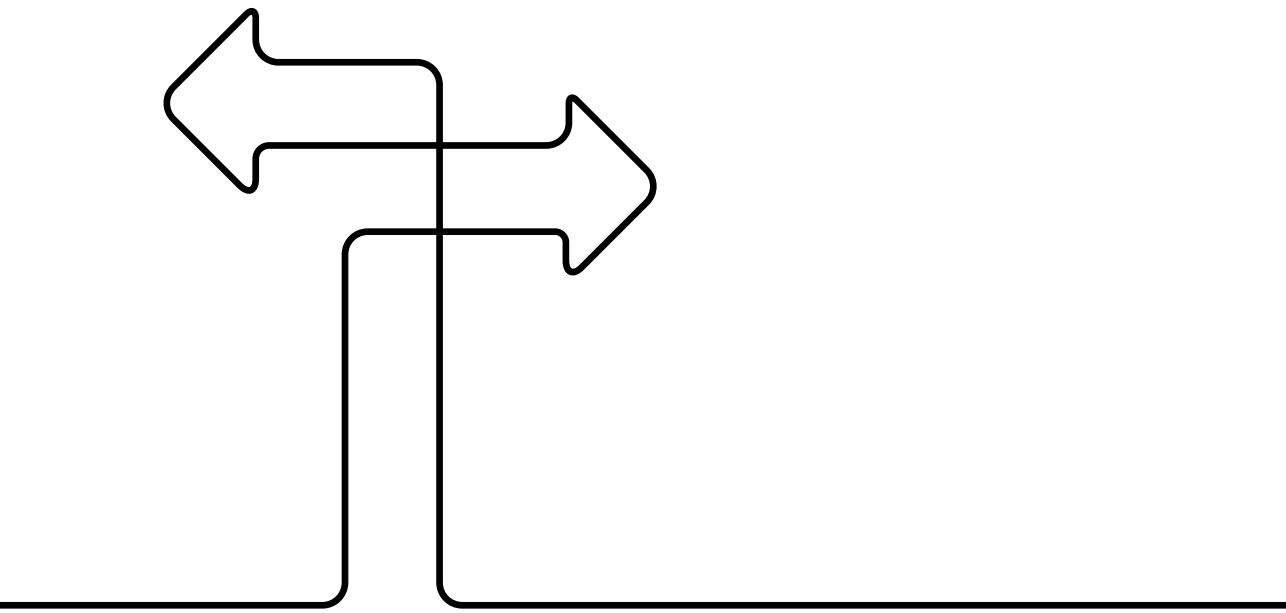
Adults

Study	TP	FP	FN	TN	Hospitalization rate	Sensitivity (95% CI)	Specificity (95% CI)
Santos 2013	485	3080	535	19515	5.0	0.48 [0.44, 0.51]	0.86 [0.86, 0.87]
Martins 2009	18265	63803	12464	222090	10.0	0.59 [0.59, 0.60]	0.78 [0.78, 0.78]
Yuksen 2016	38	68	22	392	12.0	0.63 [0.50, 0.75]	0.85 [0.82, 0.88]
Storm-Versloot 2011	66	76	96	638	18.0	0.41 [0.33, 0.49]	0.89 [0.87, 0.92]
Storm-Versloot 2011	73	72	76	651	18.0	0.49 [0.41, 0.57]	0.90 [0.88, 0.92]
Dong 2007	1519	1947	4189	21869	19.0	0.27 [0.25, 0.28]	0.92 [0.91, 0.92]
Van der Wulp 2009	2259	2146	5911	27658	21.0	0.28 [0.27, 0.29]	0.93 [0.93, 0.93]
Van der Wulp 2009	2641	3183	4415	24019	21.0	0.37 [0.36, 0.39]	0.88 [0.88, 0.89]
Fong 2018	6	5	68	221	25.0	0.08 [0.03, 0.17]	0.98 [0.95, 0.99]
Wuerz 2001	1113	674	1260	5204	28.0	0.47 [0.45, 0.49]	0.89 [0.88, 0.89]
Wuerz 2000	91	53	68	281	32.0	0.57 [0.49, 0.65]	0.84 [0.80, 0.88]
Graff 2014	6497	2132	8463	28377	33.0	0.43 [0.43, 0.44]	0.93 [0.93, 0.93]
Chi 2006	713	747	376	1336	34.0	0.65 [0.63, 0.68]	0.64 [0.62, 0.66]

Children

Study	TP	FP	FN	TN	Hospitalization rate	Sensitivity (95% CI)	Specificity (95% CI)
Al-Hindi 2014	65	295	45	2609	4.0	0.59 [0.49, 0.68]	0.90 [0.89, 0.91]
Gouin 2005	8	19	38	472	9.0	0.17 [0.08, 0.31]	0.96 [0.94, 0.98]
Gravel 2009	1937	2746	3124	50722	9.0	0.38 [0.37, 0.40]	0.95 [0.95, 0.95]
Green 2012	33	42	47	658	10.0	0.41 [0.30, 0.53]	0.94 [0.92, 0.96]
Aeimchanbanjong 2017	62	239	62	679	12.0	0.50 [0.41, 0.59]	0.74 [0.71, 0.77]
Aeimchanbanjong 2017	64	165	59	753	12.0	0.52 [0.43, 0.61]	0.82 [0.79, 0.84]
Aeimchanbanjong 2017	70	285	53	633	12.0	0.57 [0.48, 0.66]	0.69 [0.66, 0.72]
Baumann 2005	44	66	28	372	14.0	0.61 [0.49, 0.72]	0.85 [0.81, 0.88]
Chang 2013	3104	6821	6276	26327	23.0	0.33 [0.32, 0.34]	0.79 [0.79, 0.80]
Allon 2017	2631	2330	17726	60922	24.0	0.13 [0.12, 0.13]	0.96 [0.96, 0.96]

TP = true positive; FP = false positive; FN = false negative; TN = true negative



Chapter 5

Multiple performance measures are needed to evaluate triage systems in the emergency department

Joany M Zachariasse, Daan Nieboer, Rianne Oostenbrink, Henriëtte A Moll, Ewout W Steyerberg

J Clin Epidemiol. 2018 Feb;94:27-34

ABSTRACT

Objectives Emergency department triage systems can be considered prediction rules with an ordinal outcome, where different directions of misclassification have different clinical consequences. We evaluated strategies to compare the performance of triage systems and aimed to propose a set of performance measures that should be used in future studies.

Study design and setting We identified performance measures based on literature review and expert knowledge. Their properties are illustrated in a case study evaluating two triage modifications in a cohort of 14,485 pediatric emergency department visits. Strengths and weaknesses of the performance measures were systematically appraised.

Results Commonly reported performance measures are measures of statistical association (34/60 studies) and diagnostic accuracy (17/60 studies). The case study illustrates that none of the performance measures fulfills all criteria for triage evaluation. Decision curves are the performance measures with the most attractive features but require dichotomization. In addition, paired diagnostic accuracy measures can be recommended for dichotomized analysis, and the triage-weighted kappa and Nagelkerke's R^2 for ordinal analyses. Other performance measures provide limited additional information.

Conclusion When comparing modifications of triage systems, decision curves and diagnostic accuracy measures should be used in a dichotomized analysis, and the triage-weighted kappa and Nagelkerke's R^2 in an ordinal approach.

INTRODUCTION

Emergency departments face large and unpredictable numbers of patients, presenting with a broad spectrum of illnesses and injuries [3,4]. As demand often exceeds the capacity to provide immediate care, most patients have to wait before they can be seen by a healthcare professional. The vast majority of emergency care settings have triage systems in place to prioritize patients and ensure they are seen in order of clinical need rather than in order of attendance [9]. Research on triage systems is important, both to understand the performance of currently used systems, and to enable the evaluation of modifications for improvement. Nevertheless, despite the almost universal application of triage systems at the emergency departments, studies about their performance to correctly discriminate between high- and low- urgency patients are limited and hampered by methodological limitations [17, 104, 105].

Triage systems can be considered a specific type of prediction model, and their evaluation can be approached accordingly. Some important characteristics of triage systems, however, make their evaluation more challenging. Triage systems typically classify patients into five ordinal categories. There is a vast methodological literature on prediction models with dichotomous outcomes (i.e., the presence or absence of the disease), but evaluation of ordinal prediction models has been less well studied [106, 107]. Furthermore, different types of misclassification by a triage system have different clinical consequences. Classifying critically ill patients to a too low urgency level ("undertriage") leads to delays in treatment with immediate clinical consequences. Classifying nonurgent patients to a too high urgency level ("overtriage") does not have a direct effect on the patient but decreases the efficiency of the system, ultimately leading to increased waiting times for severely ill patients correctly classified as high urgent. Commonly used performance measures typically do not take into account the ordinal nature of triage systems or the different weights of the different directions of misclassification.

In this study, we aim to evaluate currently available strategies to compare the performance of triage systems. We will illustrate the challenges when assessing triage systems and the properties of several performance measures with a case study that evaluates two modifications of a commonly used triage system, one aimed to reduce overtriage and one aimed to reduce undertriage. Furthermore, we aim to propose a set of performance measures that should be used in future studies.

REVIEW OF PERFORMANCE MEASURES

We first considered performance measures based on a review of the current literature. We used the search selection from a previously conducted systematic review, including EMBASE, Medline, OvidSP, Cochrane central, Web-of-science and CINAHL databases from

1980 till 2013 to identify studies that assessed the performance of a triage system in emergency care with a predefined reference standard [60]. After an update in May 2017, a total of 60 studies were included, published between 1996 and 2017.

Of the 60 included studies, 35 (58%) used measures of statistical association to describe the performance of triage systems. The most commonly reported measures were Pearson's *chi-square test*, *t*-test, ANOVA, and their non-parametric equivalent. We also found several types of correlation coefficients (7 studies), and regression coefficients or odds ratios (9 studies). Seventeen studies (28%) reported some type of diagnostic accuracy measure, including sensitivity, specificity, predictive values, and likelihood ratios, whereas 10 reported the area under the curve, and 15 studies described some other type of performance measures such as the kappa statistic (Appendix A).

For our case study, we selected the performance measures that were used in at least two different studies, thereby excluding RIDIT analysis, univariate optimal discriminant analysis, the Kolmogorov-Smirnov test, the Wald Wolfowitz test, Friedman's test, Cox-Stuart trend test and Net reclassification improvement. In addition to these measures selected from the literature, methodologic experts (D.N. and E.W.S.) identified a number of performance measures commonly used in diagnostic and prognostic research, and several recently developed performance measures (Table 1). These included Nagelkerke's R^2, the ordinal C-statistic and decision curve analysis [106, 108, 109]. Besides the unweighted kappa, we calculated a weighted kappa using quadratic weights and a "triage-weighted kappa". The triage-weighted kappa has been proposed as an alternative weighting scheme specifically adapted to the practice of triage [110] (Appendix B).

In this article, we focus on measures of overall performance, diagnostic accuracy, discrimination, clinical usefulness, and agreement that can be used to compare triage systems' performance to correctly discriminate between high- and low-urgency patients.

We used the following criteria to evaluate the selected performance measures for their ability to evaluate and compare modifications of triage systems:

1. Can the performance measure be applied to ordinal data without requiring dichotomization?
2. (+, if the performance measure can be applied to ordinal data without dichotomization; -, if not)
3. Does the performance measure take into account the weights of different types of misclassification (overtriage or undertriage)?
4. (+, if the performance measure takes into account the different weights of both overtriage and undertriage; -, if the performance measure weighs overtriage and undertriage the same; and +/-, if the performance measure assesses over- and undertriage independently but separate)
5. Is the measure a proper scoring system: does the score improve when the statistical model fit improves?[111, 112]

Table 1. Selection of performance measures based on 60 triage studies and expert opinion

Performance measure (number of studies)	Aspect	Application in the evaluation of triage systems (case study)
Measures of association: - Pearson's chi-square test or Fishers exact test - T-test or Mann-Whitney U test - ANOVA or Kruskal-Wallis - Log-rank test (35 studies)	Association	We performed Pearson's *chi-square* tests to assess the association between the MTS and reference standard (dichotomous and ordinal)
Measures of strength of association - Regression coefficients - Risk ratios - Odds ratios (9 studies)	Association	We fitted binary logistic (dichotomous) and ordinal (ordinal) regression models with MTS urgency as independent variable and the reference standard as the outcome. We present the odds ratios for each of the different urgency categories in which the low urgency is the reference category
Correlation coefficients: - Pearson's r - Spearman's rho - Kendall's tau (7 studies)	Correlation	We assessed the correlation between the MTS and reference standard by Spearman's and Kendall's correlation coefficient because our data consists of categorical variables (dichotomous and ordinal)
Paired diagnostic accuracy measures: - Sensitivity and specificity or - Positive and negative predictive values - Positive and negative likelihood ratios (17 studies)	Discrimination	We calculated sensitivity, specificity, the positive and negative predictive values, and positive and negative likelihood ratios based on a 2x2 classification table of the MTS and reference standard (dichotomous)
Diagnostic odds ratio (4 studies)	Discrimination	We calculated the diagnostic odds ratio by dividing the positive and negative likelihood ratios that were calculated based on a 2x2 classification table of the MTS and reference standard (dichotomous)
Area under the curve (10 studies)	Discrimination	The area under the curve was calculated based on a logistic regression model with MTS urgency as independent variable and the reference standard as the outcome (dichotomous)
Ordinal c-statistic (0 study)	Discrimination	The ordinal c-statistic was calculated based on an ordinal regression model with MTS urgency as independent variable and the reference standard as the outcome (ordinal)
Nagelkerke's R^2 (1 study)	Overall performance	Nagelkerke's R^2 was calculated, based on binary logistic (dichotomous) and ordinal (ordinal) regression models with MTS urgency as independent variable and the reference standard as the outcome.
Decision curve analysis (0 studies)	Clinical usefulness	We plotted a decision curve comparing the MTS original with the MTS modifications (dichotomous)
Accuracy (2 studies)	Agreement	We calculated the accuracy based on a 3x3 classification table of the MTS and reference standard (ordinal)
Kappa statistic (2 studies)	Agreement	We calculated the unweighted kappa, a weighted kappa using quadratic weights, and a "triage-weighted kappa", proposed in one study. The triage-weighted kappa uses an alternative weighting scheme adapted to what is considered acceptable in clinical practice. We adjusted the weights to enable calculating the kappa for a 3x3 table, but we also analyzed the triage system and reference standard as a 5x5 table, because the triage-weighted kappa was developed for 5-category systems.

Abbreviations: MTS, Manchester Triage System

6. (+, if yes; -, if not)
7. Are differences in performance a measure of clinical utility?
8. (+, if a difference in two values have a meaning that can be directly related to clinical utility, in terms of better or worse patient outcome; -, if not)

CASE STUDY

We applied the selected performance measures in a case study, based on data of an ongoing project aimed at improving the Manchester Triage System (MTS) for children [31, 34, 35, 113]. For this project, we prospectively enrolled children attending the emergency department of the Erasmus MC- Sophia Children's Hospital, Rotterdam, The Netherlands. All consecutive children aged less than 16 years, who attended the emergency department between August 2009 and December 2011 with complete data (n=14,485) were included. Routinely documented data on patient demographics, triage characteristics and disposition, were extracted from the standardized electronic patient records in the hospital information system.

Manchester Triage System and modifications

The MTS is the most commonly used triage system in European emergency departments [14, 31, 35]. It is an algorithm consisting of 52 flowcharts, covering the range of presenting problems in the emergency room. Each flowchart consists of signs and symptoms called discriminators that assign patients to one of five urgency categories. Each urgency category corresponds to a maximum waiting time the patient is allowed to wait before being seen by a physician. In the Erasmus MC, a computerized version of the official Dutch translation of the MTS is used, with previously validated modifications for children with fever [35, 37].

Triage systems can be directly compared in a study where different triage systems are applied to the same patient population. Alternatively, existing triage systems can be improved through modifications that allocate specific patient categories to a higher or lower urgency category. These modifications can be "simulated" and the resulting classification of original and modified systems can be compared. In our case study, we studied two modifications of the MTS. The "fever modification" has previously been developed to reduce overtriage in children with fever [35]. In the original second edition of the MTS, all children with fever are considered high urgent and are allocated to MTS category 2, even though a large proportion of these patients present with self-limiting illnesses. The fever modification allocates children with fever aged greater than 3 months to a lower urgency category, mostly MTS urgency 3, unless other high-urgent discriminators are present. Because this modification, MTS fever, has already been implemented at the emergency department of the Erasmus MC, we assigned all children with fever to MTS urgency level 2 to obtain the original MTS.

The second modification, the "comorbidity modification" aims to reduce undertriage in children with an underlying chronic condition. Comorbidity is a known risk factor for undertriage in children in general and specifically in children with infectious symptoms [34]. The MTS does not generally consider comorbidity as a discriminator, and in children with a chronic condition, urgency classification is usually based on other signs and symptoms. Therefore, in the comorbidity modification, we assumed that all children with a complex chronic condition and infectious symptoms were assigned to MTS urgency level 2, in addition to the modifications that were implemented in the MTS fever. We compared this modification, MTS comorbidity, with the original MTS.

Statistical analysis

To determine the performance of a triage system, a reference standard as proxy for true patient urgency was required. We used a five-category combined reference standard that has been used in previous studies [31, 35]. This reference standard appraises patient urgency based on a combination of vital signs at presentation, potentially life threatening conditions, diagnostic resources, therapeutic interventions and patient disposition (Appendix C). It was generated for each patient independently of triage category. Both the reference standard and the MTS were analyzed as a dichotomous and as an ordinal variable. Dichotomization was as high urgent, MTS 1 and 2 and low urgent, MTS 3-5. This cut-off is most commonly used in literature and clinical practice [31, 43, 114]. For the ordinal approach, we categorized the MTS and the reference standard into three categories: high urgent (MTS 1 and 2), urgent (MTS 3) and low urgent (MTS 4 and 5). This categorization was based on clinical relevance and motivated by relatively low numbers in triage categories 1 and 5.

Details on the application of the different performance measures are presented in Table 1. Analyses were performed using R software, Version 3.2.0 [115]. The programming syntax is provided in Appendix D.

RESULTS OF CASE STUDY

Of all 14,485 children attending the emergency department, the original MTS classified 3,559 (24.6%) as high urgent (MTS category 1 or 2). The fever modification reclassified 1,365 patients (9.4%) into a lower MTS urgency category compared to the original MTS. Of these, the vast majority (94%) were placed one level lower, and 6% were placed two levels lower. The comorbidity modification reclassified 516 patients (3.6%) into a higher MTS urgency category compared to the original MTS. More than half of these reclassifications were changes among one urgency level, but 243 patients (47%) were reclassified two or more categories (Appendix E).

MTS fever: reducing overtriage

In the dichotomous approach (Table 2), the diagnostic accuracy measures indicated that the fever modification, aimed to reduce overtriage, achieved its goal: it improved the specificity, positive predictive value, and positive likelihood ratio of the MTS. This came at the cost of a small decrease in sensitivity and a poorer negative likelihood ratio. Most performance measures indicated that the fever modification improved the performance, although the area under the curve was the same for the original and the modified MTS. The decision curve showed a higher net benefit of the original MTS compared to the MTS fever up to a threshold for urgency of approximately 8%. The MTS fever performed better at higher thresholds (Fig. 1). The urgency of 8% reflects a situation where overtriage of 92 patients is accepted for eight true urgency classifications. In other words, one considers missing a truly urgent patient as 12 times (0.92/0.08) worse than overtriaging a non-urgent patient. The left side of the graph (threshold < 8%) reflects a clinical situation in which one is willing to accept more overtriage to identify one true urgent patient. In this situation, the original MTS should be preferred. The right side of the graph represents the opposite situation in which one prefers a triage system with less overtriage, and the MTS fever should be used.

In the ordinal approach (Table 3), the ordinal C-statistic and Nagelkerke's R^2 were clearly better for the MTS fever. The fever modification also increased the accuracy for the weighted kappa statistics, whereas no difference was found for the unweighted kappa.

All other performance measures (Appendix F) also indicated an improved performance of the MTS fever, both in the dichotomous and ordinal approach.

MTS comorbidity: reducing undertriage

In the dichotomous approach (Table 2), the diagnostic accuracy measures indicated that the comorbidity modification, aimed to decrease undertriage, improved sensitivity, negative predictive value, and negative likelihood ratio, at the cost of a lower specificity, positive predictive value and positive likelihood ratio. The area under the curve and the odds ratios both indicated a higher performance.

The decision curve showed a higher net benefit for the MTS comorbidity compared to the MTS original up to a threshold of approximately 0.09 (Fig. 1). This threshold indicates a clinical situation where missing a truly urgent patient is 10 times (0.91/0.09) worse than overtriaging a nonurgent patient. If one considers missing a truly urgent patient more than 10 times worse, the threshold probability is lower than 0.09, and the MTS comorbidity should be preferred.

In the ordinal approach (Table 3), all performance measures showed a decreased performance of the comorbidity modification.

The results of the other performance measures (Appendix F), indicated that the MTS comorbidity had a decreased performance compared with the MTS original. Only the odds ratio in the dichotomous approach indicated an improved performance, in line with the diagnostic odds ratio (Table 3).

Table 2. Results of the dichotomous approach comparing the MTS original with the MTS modifications

	MTS original	MTS fever	MTS comorbidity
Sensitivity (95% CI)	0.73 (0.70-0.76)	0.64 (0.61-0.67)	0.77 (0.75-0.80)
Specificity (95% CI)	0.79 (0.79-0.80)	0.89 (0.88-0.89)	0.76 (0.75-0.77)
Pos. predictive value (95% CI)	0.23 (0.22-0.24)	0.32 (0.30-0.34)	0.21 (0.20-0.22)
Neg. predictive value (95% CI)	0.97 (0.97-0.98)	0.97 (0.96-0.97)	0.98 (0.97-0.98)
Pos. likelihood ratio (95% CI)	3.56 (3.4-3.74)	5.74 (5.38-6.13)	3.22 (3.08-3.36)
Neg. likelihood ratio (95% CI)	0.34 (0.31-0.37)	0.41 (0.38-0.44)	0.30 (0.27-0.33)
Diagnostic Odds Ratio (95% CI)	10.56 (9.19-12.13)	14.08 (12.32-16.09)	10.81 (9.35-12.51)
Area under the curve (95% CI)	0.763 (0.750-0.777)	0.763 (0.749-0.778)	0.767 (0.754-0.780)
Nagelkerke's R^2	0.203	0.241	0.202

Abbreviations: MTS, Manchester Triage System; CI, confidence interval

Fig. 1. Decision curve, comparing the net benefit of the MTS original (Original) with the MTS fever (Fever) and the MTS comorbidity (Comorbidity)

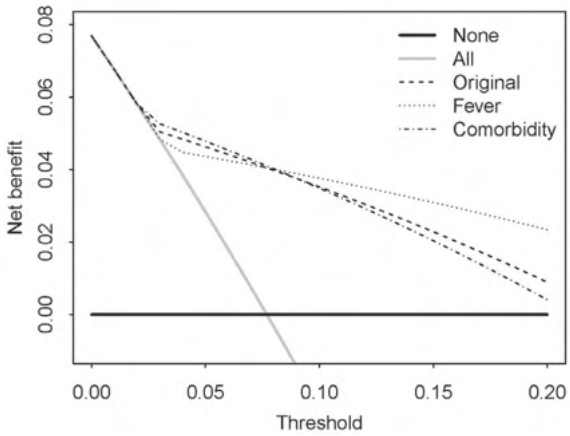


Table 3. Results ordinal approach comparing the MTS original with the MTS modifications

Performance measure	MTS original	MTS fever	MTS comorbidity
Accuracy	0.51	0.53	0.50
Ordinal c-statistic (95% CI)	0.721 (0.711-0.731)	0.737 (0.727-0.746)	0.719 (0.710-0.728)
Nagelkerke's R^2	0.126	0.179	0.123
Kappa (95% CI) *Analysed by keeping MTS and reference standard 3x3*			
Unweighted	0.23 (0.21-0.24)	0.23 (0.22-0.24)	0.22 (0.21-0.23)
Weighted (quadratic)	0.29 (0.27-0.30)	0.36 (0.35-0.37)	0.27 (0.26-0.28)
Triage-weighted	0.23 (0.22-0.24)	0.24 (0.23-0.26)	0.22 (0.21-0.23)
Kappa (95% CI) *Analysed by keeping MTS and reference standard 5x5*			
Unweighted	0.14 (0.13-0.15)	0.14 (0.13-0.15)	0.13 (0.12-0.14)
Weighted (quadratic)	0.34 (0.33-0.36)	0.39 (0.38-0.41)	0.33 (0.32-0.35)
Triage-weighted	0.22 (0.21-0.23)	0.26 (0.25-0.27)	0.21 (0.20-0.22)

Abbreviations: MTS, Manchester Triage System; CI, confidence interval

APPRAISAL OF PERFORMANCE MEASURES

The results of the evaluation of the selected performance measures are presented in Table 4. None of the selected performance measures fulfilled all criteria. Decision curves and the paired diagnostic accuracy measures are the performance measures with the most attractive features, but require dichotomization. The (triage-)weighted kappa and Nagelkerke's R^2 are the most favorable performance measure that can be applied to ordinal data.

Table 4. Appraisal of different performance measures

Performance measure	Ordinal[1]	Weights[2]	Proper scoring system[3]	Interpretation of differences[4]
Sensitivity and specificity Positive and negative predictive values Positive and negative likelihood ratios	-	+/-	-	+
Diagnostic odds ratio	-	-	-	-
Accuracy	+	-	-	+
Area under the curve and Ordinal c-statistic	+	-	-	-
Nagelkerke's R^2	+	-	+	-
Decision curve analysis	-	+	+	+
Kappa statistic				
Unweighted	+	-	-	-
Weighted (quadratic)	+	-	-	-
Triage-weighted	+	+	-	-

[1] Can the performance measure be applied to ordinal data?
[2] Does the performance measure take into account the weights of different types of misclassification?
[3] Is the measure a proper scoring system?
[4] Are differences in performance measure a measure of clinical utility?

Proposed performance measures for the evaluation of triage systems

The diagnostic accuracy measures quantify performance of a triage system separately for both directions of misclassification. In the case study, the diagnostic accuracy measures indicated how the modifications achieved their distinct purposes. The fever modification, aimed to reduce overtriage, improved the specificity, positive predictive value, and positive likelihood ratio of the MTS at the cost of a slightly poorer sensitivity, negative predictive value, and negative likelihood ratio. The opposite was true for the comorbidity modification that aimed to reduce undertriage. Diagnostic accuracy measures are commonly used, and their results have a direct and intuitive clinical interpretation. However, they do not "weigh" the different directions of misclassification and therefore do not indicate what modification should be preferred. It remains, for example, unclear what higher sensitivity is accepted at the cost of a certain lower specificity.

The decision curve is the only measure that takes into account the different weights of different types of misclassification with a direct clinical interpretation. Performance of a triage system is graphically displayed in a curve plotting the net benefit of a triage system over a range of threshold probabilities. Triage modifications can be directly compared in the same populations: the modification with the highest net benefit indicates the better alternative. Thresholds can be used to incorporate the uncertainty about the exact weight of the harms and benefits of over- and undertriage. Even if experts do not agree on the exact threshold, it is likely that they agree on a range of thresholds that is relevant for clinical practice. Moreover, different ranges of thresholds can be used to represent different clinical situations.

In the case study, the decision curves indicated that the MTS comorbidity should be preferred up to a threshold of approximately 0.09, whereas the MTS fever should be favored in settings with a threshold probability of 0.08 or higher. This suggests that the comorbidity modification is the preferred modification if one accepts to find one seriously ill patient at the cost of 10 or more overtriaged patients. In settings where this thresholds is higher**,** the harm of false-positives receives more weight and the fever modification is preferred.

A major drawback of the diagnostic accuracy measures and decision curve analysis is that they require dichotomization of the data, whereas the triage systems are constructed as ordinal systems. When comparing the performance of triage systems with their potential modifications this does not matter too much if the modifications involve reclassifications across the cutoff that is used for dichotomization. In our case study, we evaluated two modifications that predominantly involved reclassifications between the high urgent patients (MTS 1 and 2) and the middle- and low-urgency categories (MTS 3-5), which was also our cutoff for dichotomization. In case the modifications involve reclassifications across other urgency levels, e.g., between high or middle (MTS 1-3) vs. low urgency levels (MTS 4 and 5), a different cutoff for dichotomization can be used, and diagnostic accuracy measures and decision curves can be provided in the same way. However, when modifications involve reclassifications across multiple categories, such dichotomization will lead to a loss of information.

The kappa statistic can be used to measure the degree of agreement between a triage system and an ordinal reference standard, adjusted for possible agreement occurring by chance. The unweighted kappa only gives credit for full agreement between triage system and reference standard. Overtriage and undertriage are essentially weigthed as “0” and therefore do not contribute to the statistic. Weighted kappa on the other hand give credit for partial agreement between triage level and reference standard by assigning different weights. The commonly used quadratic weighting scheme assigns less weight to categories that are further apart but does not take into account the direction of misclassification. The triage-weighted kappa uses an asymmetric weighting scheme that gives less credit to undertriage than to overtriage. This scheme was developed by experts to reflect the clinical practice of triage.

Different experts may favor different weights and weights may depend on the study population or setting. Therefore, it can be argued that the kappa's weights are not completely objective. Moreover, weights do not have a "unit" and consequently do not have a direct clinical interpretation, unlike the net benefit that is used in the decision curve. Therefore, we would argue that decision curves should be preferred in case of a dichotomous analysis. The triage-weighted kappa should be used in an ordinal evaluation: it takes into account the different weights of over- and undertriage in a single meaure, thus enabling comparison between triage modifications.

In our case study, we observed that the absolute values of the kappa statistic were low, due to an imbalance in the groups. Only 8% of patients were classified as high urgent according to our reference standard. This is a common scenario in the emergency department where low-urgency patients are generally the majority. It will therefore be difficult to interpret the absolute values of kappa according to the rule of thumb by Landis and Koch.

Nagelkerke's R^2 is the second performance measure with favorable properties that can be applied to ordinal data. It summarizes performance in a single measure, thereby enabling a direct comparison between two triage systems. A limitation of this performance measure is that it implicitly gives the same weights to the different directions of misclassification. Consequently, a higher R^2 should not be automatically interpreted as "better" if the direction of the improvement is not known.

Other performance measures

Accuracy is a commonly used performance measure that can be applied to ordinal data. As a descriptive statistic, it may provide some insight in the effect of modifications, especially if the proportions of over- and undertriage are reported as well. A major limitation is that in unbalanced datasets, it does not reflect prediction performance for the smaller class. Because the triage systems are designed to identify the relatively small group of high-urgency patients among the larger group of lower urgency patients, accuracy cannot be proposed as a recommended performance measure.

The diagnostic odds ratio and the ordinal C-statistic, an extension of the area under the curve, are additional performance measures that can be applied to ordinal data. In our case study, the AUC, diagnostic odds ratio, and Nagelkerke's R^2 gave inconsistent results. Because the AUC and diagnostic odds ratio are not proper scoring rules, they may not be optimal for the best triage system. Therefore, their results might be different from a proper scoring rule, such as Nagelkerke's R^2.

Measures of association and correlation are frequently used in the literature but have many disadvantages. As they merely express the strength of the association between two variables, their values do not have a meaningful interpretation. We hence object to the practice where a triage system that shows a statistically significant association with a reference standard is interpreted as "valid" [116-118].

CONCLUSION AND RECOMMENDATIONS

Our results support the use of multiple performance measures when evaluating triage systems. Considering the strengths and weaknesses of each of the performance measures, we suggest that researchers provide at least a set of diagnostic accuracy measures and a decision curve analysis, based on dichotomization of the data using a relevant cutoff point, and calculate the triage-weighted kappa and Nagelkerke's R^2 for ordinal data.

We recognize that our study has some limitations. First, we do not provide an exhaustive overview of all available performance measures that could be used for the evaluation of triage systems. We made a selection based on a literature review and expert knowledge, with the aim to select the most relevant performance measures.

Second, the effect of the modifications on the performance of the triage system was relatively small, which makes the interpretation of differences in performance difficult. This is, however, a reflection of clinical practice. Triage systems are applied to a wide range of heterogeneous conditions, and improvements of triage systems are generally based on a specific subset of conditions or patients, as illustrated by the examples in our case study. Modifications will therefore typically lead to a relatively small proportion of reclassifications. It is important that performance measures capture small but clinically relevant differences.

Moreover, we used as a case study a cohort of consecutive emergency department visits that took place between August 2009 and 2011. However, because the features of performance measures should be independent of the sample on which they are applied, we do not believe a different study period will change our conclusion.

Future research should aim to provide performance measures for triage systems with multiple ordinal levels. Specifically, we encourage researchers to extend the work on decision curve analysis for outcomes with multiple categories. Guidelines and recommendations for clinicians and researchers aiming to improve triage systems will increase the quality and generalizability of future studies.

SUPPORTING INFORMATION

Available as online web appendix on the website of Journal of Clinical Epidemiology:
- Appendix A. Literature review
- Appendix D. Programming syntax

Appendix B. Weighted kappa

Table B.1 Quadratic weights for an ordinal scale with 3 categories

	1	**2**	**3**
1	1	0.75	0
2	0.75	1	0.75
3	0	0.75	1

Algorithm for quadratically weighted kappa: $W_{ij} = 1 - (i - j)^2 / (c - 1)^2$
i = category rated by rater "x"
j = category rated by rater "y"
c = number of categories of a certain scale

Table B.2 Quadratic weights for an ordinal scale with 5 categories

	1	**2**	**3**	**4**	**5**
1	1	0.94	0.75	0.44	0
2	0.94	1	0.94	0.75	0.44
3	0.75	0.94	1	0.94	0.75
4	0.44	0.75	0.94	1	0.94
5	0	0.44	0.75	0.94	1

Algorithm for quadratically weighted kappa: $W_{ij} = 1 - (i - j)^2 / (c - 1)^2$
i = category rated by rater "x"
j = category rated by rater "y"
c = number of categories of a certain scale

Table B.3 Triage-adjusted weights for an ordinal scale with 3 categories

	1	**2**	**3**
1	1	0.38	0
2	0.19	1	0.5
3	0	0.33	1

Algorithm for triage weighted kappa:
Overtriage: $W_{ij} = (1 - ((i - j)^2 / (c - 1)^2) * (i/j)$
Undertriage: $W_{ij} = (1 - ((i - j)^2 / (c - 1)^2) * (i/j)^2$
i = category rated by rater "x"
j = category rated by rater "y"
c = number of categories of a certain scale

Table B.4 Triage-adjusted weights for an ordinal scale with 5 categories

	1	**2**	**3**	**4**	**5**
1	1	0.47	0	0	0
2	0.23	1	0.63	0	0
3	0	0.42	1	0.70	0.45
4	0	0	0.53	1	0.75
5	0	0	0	0.60	1

Algorithm for triage weighted kappa:
Overtriage: $W_{ij} = (1 - ((i - j)^2 / (c - 1)^2) * (i/j)$
Undertriage: $W_{ij} = (1 - ((i - j)^2 / (c - 1)^2) * (i/j)^2$
i = category rated by rater "x"
j = category rated by rater "y"
c = number of categories of a certain scale

Reference

van der Wulp I, van Stel HF. Adjusting weighted kappa for severity of mistriage decreases reported reliability of emergency department triage systems: a comparative study. J Clin Epidemiol. 2009;62:1196-201

Appendix C. Reference standard

Table C.1 Reference classification matrix

	Diagnostics					Therapy			Follow-up		
	Vital	**PLC**	**Simple**	**Imaging**	**Extensive**	**Rx**	**Rx at ED**	**Intervention**	**Tel./GP**	**Outpatient**	**Hospitalization**
Emergent	1	n/a		n/a			n/a			n/a	
Very urgent	0	1		n/a			n/a			n/a	
Urgent	0	0		n/a		0	0	1		n/a	
	0	0		n/a		0	1	0	n/a	0	1
	0	0	1	0	0		n/a		n/a	0	1
	0	0	0	1	0		n/a		n/a	0	1
	0	0	0	0	1		n/a		n/a	0	1
	0	0	0	1	0	1	0	0	n/a	1	0
	0	0	0	0	1	1	0	0	n/a	1	0
	0	0	0	1	0	0	1	0		n/a	
	0	0	0	0	1	0	1	0		n/a	
	0	0	0/**1**	1	1		n/a				
Standard	All other combinations										
Non-urgent	0	0	0	0	0	0/**1**	0	0	0	0	0

1=present/0=absent; n/a = not applicable; PLC=possible life threathening condition; Rx = medication on prescription

Table C.2 Definitions of reference urgency categories

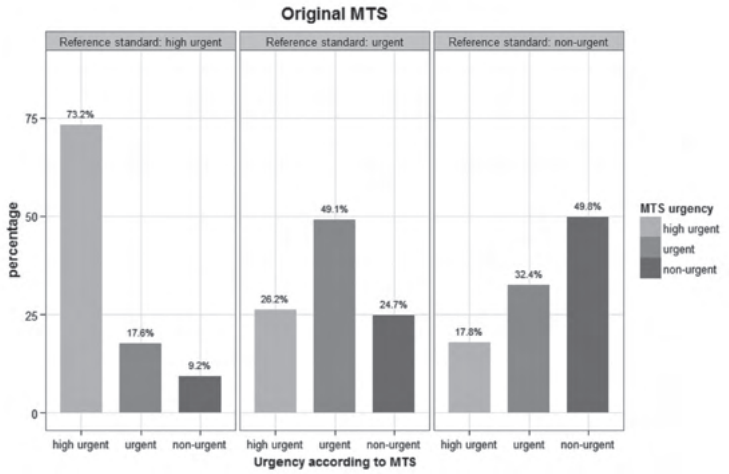
Urgency category	Definition
Immediate	Patients with abnormal vital signs according to the Pediatric Risk of Mortality Score III[2]
Very urgent	Patients diagnosed with life-threatening conditions, defined as meningitis, sepsis, high energetic trauma, substantial blood loss, aorta dissection, >10% dehydration, (near) drowning, electric trauma, possible dangerous intoxication, >10% burns, and facial burns or possible inhalation trauma
Urgent	Patients who received intravenous medication (including aerosols and fluids) or casting or inguinal hernia reposition or luxation reposition or gastrolavage at the ED; Patients who had some diagnostic workup or received oral medication or small surgical interventions, e.g. bandage at the ED and were admitted to hospital; Patients who had extended laboratory diagnostics including blood culture, cerebrospinal fluid puncture or multiple laboratory tests or imaging and who received therapy at the ED or small surgical interventions; Patients who had imaging and extended laboratory diagnostics; Patients who had extended laboratory diagnostics or imaging at the ED, received some therapy (including medication on prescription or simple advice) at the ED, and had a planned followup visit within 24 h
Standard	Patients with some diagnostic workup or therapy at the ED or were admitted to hospital or had a planned follow-up visit without meeting the criteria for urgent
Non-urgent	Patients with no diagnostic workup, no treatment at the ED, and who were discharged without a planned follow-up visit

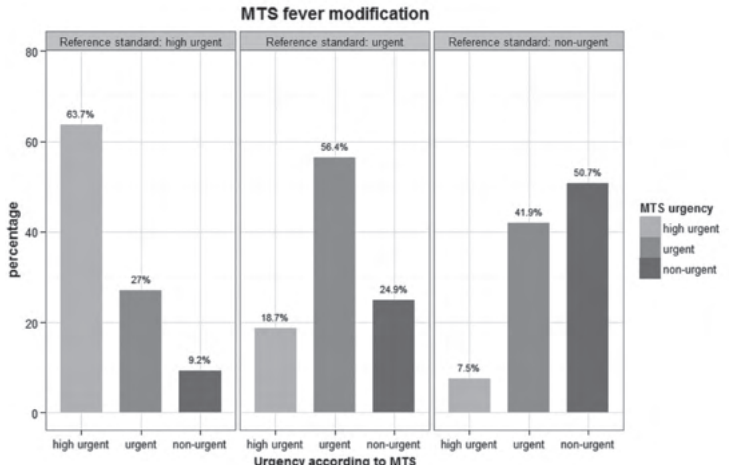
References

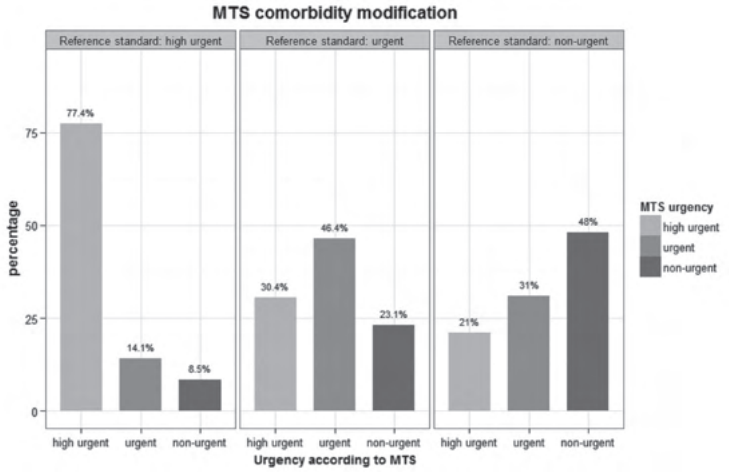
van Veen M, Steyerberg EW, Ruige M, van Meurs AH, Roukema J, van der Lei J, et al. Manchester triage system in paediatric emergency care: prospective observational study. BMJ. 2008;337:a1501.

Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. Crit Care Med. 1996;24(5):743–752

Appendix E. MTS urgencies per reference standard category







Appendix F. Results from other performance measures

Table F.1 Results of the dichotomous approach comparing the MTS original with the MTS fever

	MTS original	MTS fever
Pearson's chi-square test	$X^2(1) = 1534.5$ $p < 0.0001$	$X^2(1) = 2212.6$ $p < 0.0001$
Spearman's rho Kendall's tau	0.326 0.326	0.391 0.391
Odds ratio (95% CI)	10.56 (9.20-12.15)	14.08 (12.32-16.10)

Table F.2 Results ordinal approach comparing the MTS original with the MTS fever

	MTS original	MTS fever
Pearson's chi-square test	$X^2(4)= 2330$ $p < 0.0001$	$X^2(4)= 3120$ $p < 0.0001$
Spearman's rho Kendall's tau	0.318 0.297	0.367 0.345
Odds ratio (95% CI) MTS urgency 1+2 MTS urgency 3 MTS urgency 4+5	 5.87 (5.36-6.44) 2.85 (2.62-3.10) Reference	 12.68 (11.37-14.15) 2.73 (2.52-2.95) Reference

Table F.3 Results of the dichotomous approach comparing the MTS original with the MTS comorbidity

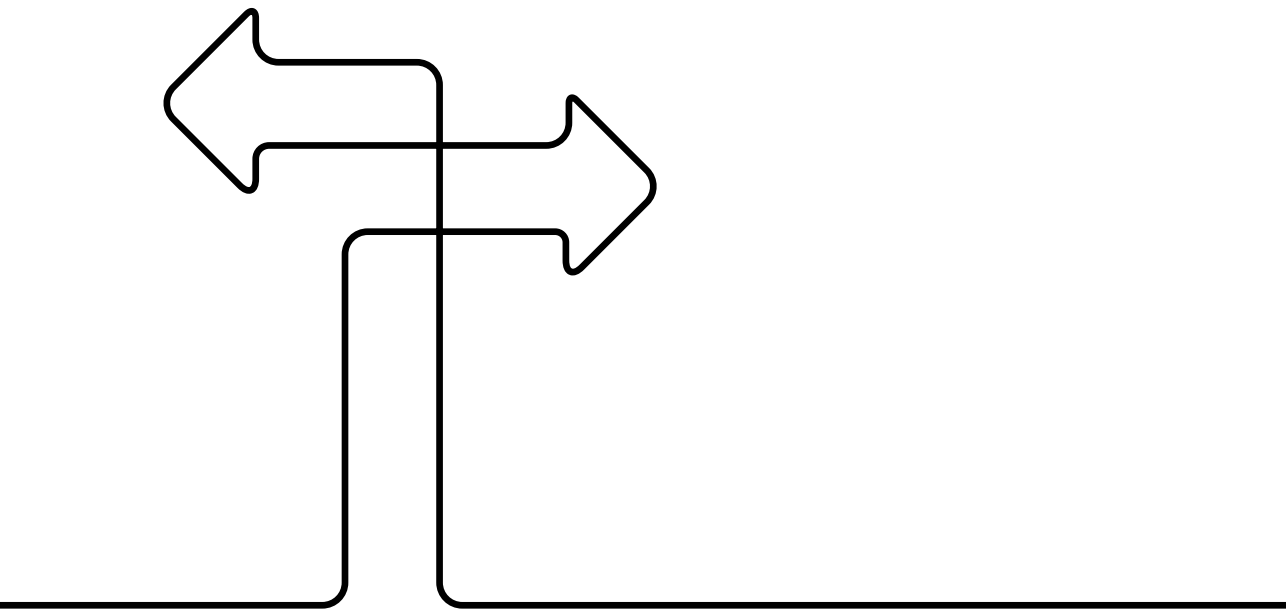
	MTS original	MTS comorbidity
Pearson's chi-square test	$X^2(1) = 1534.5$ $p < 0.0001$	$X^2(1) = 1445$ $p < 0.0001$
Spearman's rho Kendall's tau	0.326 0.326	0.316 0.316
Odds ratio (95% CI)	10.56 (9.20-12.15)	10.81 (9.36-12.53)

Table F.4 Results ordinal approach comparing the MTS original with the MTS comorbidity

	MTS original	MTS comorbidity
Pearson's chi-square test	$X^2(4) = 2330$ $p < 0.0001$	$X^2(4) = 2231$ $p < 0.0001$
Spearman's rho Kendall's tau	0.318 0.297	0.314 0.293
Odds ratio (95% CI) MTS urgency 1+2 MTS urgency 3 MTS urgency 4+5	 5.87 (5.36-6.44) 2.85 (2.62-3.10) Reference	 5.57 (5.09-6.09) 2.85 (2.61-3.10) Reference

PART II

Improving the Manchester Triage System for children



Chapter 6

Improving the prioritization of children at the emergency department: Updating the Manchester Triage System using vital signs

Joany M Zachariasse, Ian K Maconochie, Ruud Nijman, Susanne Greber-Platzer, Frank J Smit, Daan Nieboer, Johan van der Lei, Claudio F Alves, Henriëtte A Moll

Resubmitted to PLOS One with Minor Revisions

ABSTRACT

Background Vital signs are used in emergency care settings in the first assessment of children to identify those that need immediate attention. We aimed to develop and validate vital sign based Manchester Triage System (MTS) discriminators to improve triage of children at the emergency department.

Methods and findings The TrIAGE project is a prospective observational study based on electronic health record data from five European EDs (Netherlands (n=2), United Kingdom, Austria, and Portugal). In the current study, we included 117,438 consecutive children <16 years presenting to the ED during the study period (2012-2015). We derived new discriminators based on heart rate, respiratory rate, and/or capillary refill time for specific subgroups of MTS flowcharts. Moreover, we determined the optimal cut-off value for each vital sign. The main outcome measure was a previously developed 3-category reference standard (high, intermediate, low urgency) for the required urgency of care, based on mortality at the ED, immediate lifesaving interventions, disposition and resource use. We determined six new discriminators for children <1 year and ≥1 year: "Very abnormal respiratory rate", "Abnormal heart rate", and "Abnormal respiratory rate", with optimal cut-offs, and specific subgroups of flowcharts. Application of the modified MTS reclassified 744 patients (2.5%). Sensitivity increased from 0.66 (95%CI 0.60-0.72) to 0.71 (0.66-0.75) for high urgency patients and from 0.67 (0.54-0.76) to 0.70 (0.58-0.80) for high and intermediate urgency patients. Specificity decreased from 0.90 (0.86-0.93) to 0.89 (0.85-0.92) for high and 0.66 (0.52-0.78) to 0.63 (0.50-0.75) for high and intermediate urgency patients. These differences were statistically significant. Overall performance improved (R2 0.199 versus 0.204).

Conclusions Six new discriminators based on vital signs lead to a small but relevant increase in performance and should be implemented in the MTS.

INTRODUCTION

Triage is a quick assessment to prioritize patients upon presentation to the emergency department (ED), according to the acuity of their presenting condition. In Europe, the Manchester Triage System (MTS) is the most widely used emergency medical triage system for the triage of adults and children.[14] Previous research has shown that validity of the MTS is moderate to good, with lowest performance in children and elderly.[13, 119] In a recent large prospective study in three European hospitals, sensitivity of the MTS in children ranged from 0.65 (95%CI 0.61-0.70) to 0.83 (95%CI 0.79-0.87), and specificity from 0.83 (95%CI 0.82-0.83) to 0.89 (95%CI 0.88-0.90).[119] Improvement of the MTS and particularly it's sensitivity is needed to improve the correct identification of seriously ill children and avoid harm by delays in care.[23]

Physiological parameters have been shown early markers of patient deterioration in hospital wards.[18-20] Moreover, children with severe undertriage often have abnormal vital signs.[23] Certain vital parameters, such as oxygen saturation and temperature are integrated within the flowcharts of the MTS, and severe deviations such as airway compromise and shock are included in a descriptive manner. The MTS, however, does not require routine measurement of heart rate, respiratory rate, and capillary refill time, although these vital signs are considered important predictors of severe disease.[102, 120-123] A previous study evaluated the addition of heart rate and respiratory rate to the MTS, but concluded that the use of vital signs did not improve MTS performance.[113] This study, however, added vital signs to all flowcharts in the MTS, applied pre-defined cut-offs, and used hospitalization as the reference standard. We hypothesize that with a different study design and better reference standard, an improvement in performance may be achieved.

The current study aims to develop and validate modifications to the MTS based on vital signs to improve the triage of children at the ED. This study explores the added value of vital signs to specific flowcharts, with optimal cut-off values, using a 3-category reference standard that is a proxy for true patient urgency.

METHODS

The current study was embedded in the TrIAGE project, a European prospective observational study, based on electronic health record data. The study was approved by the participating institutions' medical ethical committees: Medical Ethics Committee Erasmus MC (MEC-2013-567), Maasstad Ziekenhuis Board of Directors (Protocol L2013-103), Imperial College London Joint Research Compliance Office (Reference number: 14SM2164; Ethics reference number 14/WA/1051), Comissão de Ética para a Saúde do Hospital Prof. Dr. Fernando Fonseca EPE (Reunião de 06 de Dezembro de 2017), Ethik Kommission Medizinische der Medizinischen Unversität Wien (EK Nr: 1405/2014). All

waived the requirement for informed consent. We followed the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) statement for reporting (S6 File).[124]

Settings, study population and data collection

The TriAGE project is described in detail elsewhere.[125] In short, a cohort was established, consisting of all consecutive ED-visits of children under the age of 16 years. Participating study sites included five EDs in the Netherlands (n=2), United Kingdom, Austria and Portugal. Enrolment took place during a study period of 8 to 36 months between 2012 and 2015. Nurses routinely recorded patient characteristics, triage details, vital signs and patient disposition in each hospitals' electronic health record system. These data were automatically extracted, harmonized and checked for quality. The TrIAGE study was based on a convenience sample from five diverse ED settings. Based on projections from the participating hospitals and a pilot in the Erasmus MC, the study was designed to include at least 100 high urgency patients per hospital and at least 100 high urgency patients in the ten most commonly used MTS flowcharts.

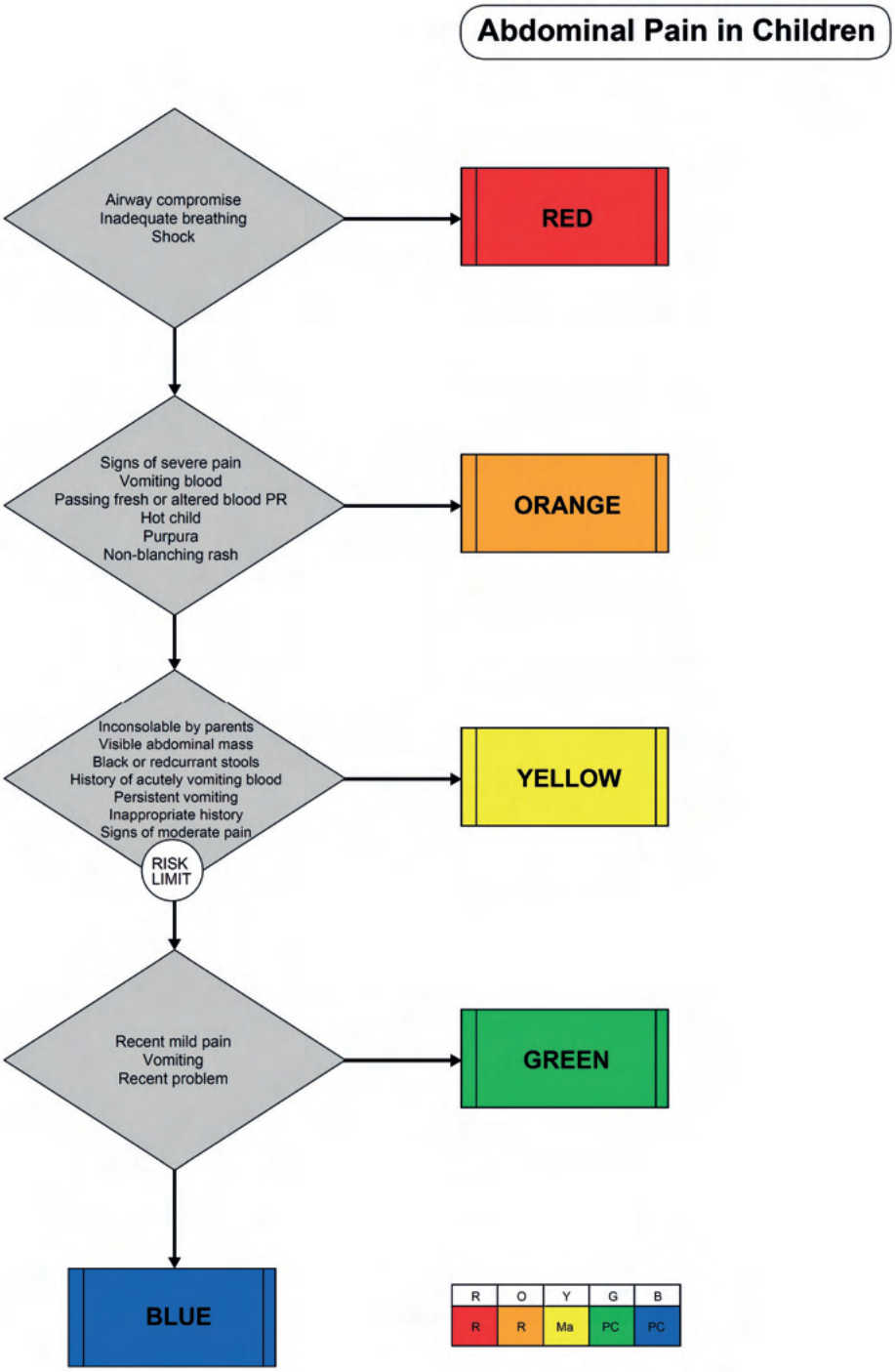
Manchester Triage System

The MTS is a flowchart-based emergency medical triage system. It consists of 52 flowcharts that cover almost all presenting problems in the ED. Flowcharts in turn consist of additional signs and symptoms named discriminators that discriminate between five clinical priorities (Immediate, Very urgent, Urgent, Standard or Non-urgent) (Fig 1). Each urgency level has been given a maximum waiting time before first contact with the treating clinician, ranging from 0 minutes (Immediate) to 240 minutes maximum waiting time (Non-urgent) Because of the low proportion of patients in the Immediate (0.8%) and the Non-urgent category (1.4%), we combined the categories Immediate and Very urgent in a high urgency category, and the categories Standard and Non-urgent in a low urgency category for the analysis. For the current study, we excluded all patients with missing MTS urgency or MTS flowchart.

Vital Signs

To improve the MTS, we assessed the value of the vital signs heart rate, respiratory rate and capillary refill time. Other physiological measurements were already included in the MTS (consciousness, temperature, oxygen saturation and increased work of breathing) or not routinely measured at the participating EDs (blood pressure). Vital signs were measured according to each ED's local practice. Heart rate was measured using a monitor device and respiratory rate was measured manually. Capillary refill time was measured by either pressing the sternum (central CRT) or fingertip (peripheral CRT) for 5 seconds, and >2 seconds was defined as abnormal.[102] We imputed missing values 25 times using a multiple imputation model including predictors, outcome and relevant case-mix variables (S1 File).[126] The analyses were performed 25 times and pooled for a final result.

Fig 1. Example MTS flowchart



Reference standard

A predefined, composite reference standard was developed for assigning patients to a high, intermediate or low urgency category and serves as a proxy for each child's true urgency (S1 Table).[125] Items are based on information from the entire ED visit, including resource use, immediate lifesaving interventions, disposition, and mortality at the ED. These items were selected because they are markers of patient urgency upon presentation to the ED and reflect the time a patient can be allowed to wait before first contact with a physician.

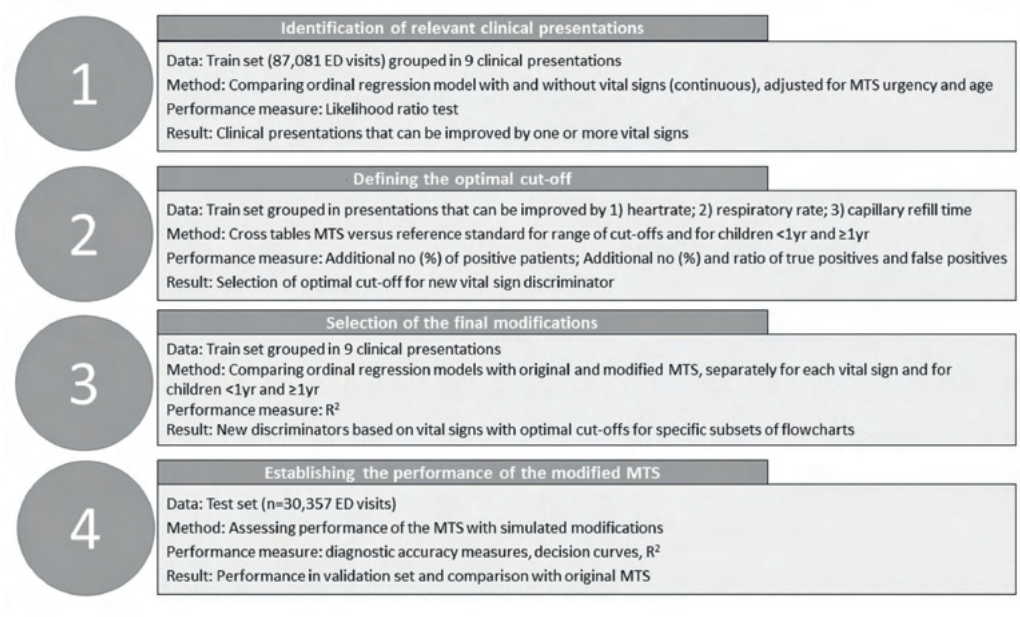
Principles for improving the MTS

To modify the MTS, one or more vital signs with a specific cut-off should be added as separate discriminator to one or more of the flowcharts. An example could be to add "Heart rate ≥120 beats per minute" as Very urgent discriminator to the flowcharts *Major Trauma* and *Wounds*. This would place all patients with a heart rate ≥120, initially triaged to one of the low urgency categories, in the Very urgent category. According to the MTS' original principles, discriminators may appear in multiple flowchart but must always lead to the same priority. Discriminators with different levels of severity, lead to different urgency levels. E.g., the MTS discriminator "Very low SaO2" (a saturation <95% on O2 therapy or <90% on air) leads to priority Very urgent, while "Low SaO2" (a saturation <95% on air) leads to priority Urgent throughout the MTS. Adding new MTS discriminators should carefully balance the safety and efficiency of any potential modification. While the goal of triage systems is to recognize patients with the highest clinical priority, it is almost as important to identify the patients with less urgent conditions. In case too many patients are falsely given a high priority, so-called "overtriage", this group may delay the diagnosis and management of the truly high-urgent patients.

Statistical analysis

Before conducting the analysis, the dataset was split into a training and a test set, based on time. The initial 75% of arrival dates per setting were assigned to the training set and were used to derive and assess the potential modifications of the MTS. The last 25% of visits were assigned to the test set and used to determine the performance of the modified MTS. To statistically and systematically assess the benefit of adding vital signs as a discriminator, and to take into account the aforementioned principled we approached the analysis in four steps (Fig 2). Details on the methodology can be found in the supplement (S2 File).

Fig 2. Schematic overview of the methodological approach



Step 1: Identification of subgroups of MTS flowcharts where a novel vital sign discriminator could have the potential to improve triage

We grouped all MTS flowcharts into nine clinical presentations: Cardiac, Dermatological, Ear Nose Throat, Gastrointestinal, Neurologic or Psychiatric, Respiratory, Trauma or Muscular, General malaise, Uro- or gynaecological and Other, according to a previous study.[119] For each clinical presentation we evaluated an ordinal regression model including MTS urgency level, age (<1 year and ≥1 year), heart rate, respiratory rate and capillary refill time as predictors and the 3-category reference standard as the outcome. Heart rate and respiratory rate were maintained as continuous variables in the analysis. This model assesses the association between vital signs and our outcome measures, adjusted for the already given MTS classification. We applied the likelihood ratio test and considered a p-value of <0.05 statistically significant. As a result, three partly overlapping groups were identified: subgroups of MTS flowcharts that could potentially be improved with the addition of a heart rate-discriminator, flowcharts that could potentially be improved with a respiratory rate discriminator and flowcharts that could potentially be improved with a capillary refill time discriminator.

Step 2: Determination of each vital sign's optimal cut-off to develop definitions for the new vital sign discriminators

For the continuous vital signs heart rate and respiratory rate, a cut-off for both high and intermediate urgency was needed. To determine the optimal cut-off value, we calculated

cross tabulations showing the association between the dichotomized MTS (high vs low triage classification) and the dichotomised reference standard (high vs low urgency). We simulated multiple triage modification where a vital sign above a certain cut-off value would place patients in the high urgency triage level. This process was repeated for the range of relevant cut-off values, with increasing steps of 10 beats per minute for heat rate and 5 breaths per minute for respiratory rate. We selected the optimal cut-off value according to three principles, based on consensus from the research team. First, we considered only thresholds with a maximum of 20% increase in the total number of positive patients. Second, we limited the ratio additional true positives : additional false positives, based on consensus from the TrIAGE research group. For the high urgency discriminators, we found a ratio 1 true positive : 15 false positives the maximum acceptable, for the high to intermediate urgency categories this was 1 true positives : 10 false positives. If multiple cut-off values were appropriate, we selected the cut-off with the largest increase in true positive patients.

Step 3: Selection of discriminators and subgroups of MTS flowcharts where performance is improved, to determine the final set of modifications

We evaluated ordinal regression models assessing the association between the original MTS and the reference standard and subsequently between the modified MTS and the reference standard. We performed analyses for each of the new discriminators separately. We selected models with a performance that was better than the original MTS as defined by a higher R2.

Step 4: Assessment of the modified MTS' performance as compared with the original MTS

In the test set, we assessed the performance of the modified MTS, i.e. the MTS with the new vital signs discriminators, as compared with the original MTS. We applied the diagnostic accuracy measures sensitivity, specificity and likelihood ratios, constructed decision curves, and calculated Nagelkerke's R2 in an ordinal analysis. These measures were selected based on our previous study evaluating performance measures in the assessment of modifications for triage systems.[127] Diagnostic accuracy measures were calculated for each of the hospitals individually, and pooled using a random effects model. To determine statistical significance, we used bootstrapping to calculate the differences between sensitivity and specificity of the original and modified MTS in a random sample with replacement and repeated this process 1000 times. We calculated the bootstrapped confidence intervals and p-value. Decision curves provide additional information about clinical value of the proposed modification by incorporating the trade-off between over- and undertriage.[109] Given that most EDs have limited capacity to see all high urgency patients at the same time, and based on consensus from the research group, we do not consider overtriage of more than nine patients acceptable in order to find one true high urgency patient.

To explore, the impact of the multiple imputation on the results, we conducted a sensitivity analysis in the original dataset with missing values. In this analysis, we assumed that the missing vital signs were not considered in the triage decision. All analyses were performed in R version 3.6.1.

RESULTS

The TrIAGE study is based on a cohort 119,209 ED visits. 1,771 children (1.5%) were excluded due to missing MTS urgency or MTS flowchart leaving a study population of 117,438 children. According to the reference standard 2,964 children (2.5%) were classified as high, 27,826 (24%) as intermediate and 86,648 (74%) as low urgent. The training set consisted of 87,081 children (74%) and the test set of 30,357 (26%) (Table 1). Heart rate was reported in 58% of children, respiratory rate in 48% and capillary refill time in 49% (S1 File).

Table 1. Baseline characteristics of the study population

	Training set (n=87,081, 74%)	**Test set (n=30,357, 26%)**
Age, no. (%)		
< 1 year	13,561 (16)	4,758 (16)
≥ 1 year	73,520 (84)	25,599 (84)
Sex, no. (%)		
Female	39,880 (46)	14,173 (47)
MTS urgency, no. (%)		
Immediate | Very urgent	9,715 (11)	3,747 (12)
Urgent	23,824 (27)	8,138 (27)
Standard | Non-urgent	53,542 (61)	18,472 (61)
Presenting problem, no (%)		
Cardiac	1,010 (1)	388 (1)
Dermatological	11,535 (13)	3,087 (10)
Ear, Nose and Throat	8,215 (9)	3,347 (11)
Gastrointestinal	13,686 (16)	4,531 (15)
Neurologic or psychiatric	3,441 (4)	1,250 (4)
Respiratory	9,640 (11)	4,320 (14)
Trauma or musculoskeletal	16,274 (19)	5,039 (17)
General malaise	7,402 (9)	2,615 (9)
Uro- or gynaecological	1,961 (2)	622 (2)
Other	13,917 (16)	5,158 (17)
Disposition, no. (%)		
ICU or mortality at ED*	467 (0.5)	217 (0.7)
Hospital admission	8,436 (10)	3,004 (10)
Discharge / other	78,178 (89)	27,136 (89)

*Mortality: 12 patients in train set and 4 in test set

New vital signs discriminators

Heart rate showed a significant association with the reference standard in the cardiac, dermatological, neurologic/psychiatric, and respiratory clinical presentations. For the discriminator very abnormal heart rate (high urgency triage level), we selected as optimal cut-off a heart rate ≥170 beats per minute for children <1 year, and for the discriminator abnormal heartrate (intermediate urgency), ≥160 beats per minute for children <1 year and ≥140 beats per minute for children ≥1 year. A cut-off for children ≥1 year could not be defined, because all of the explored cut-off values led to inacceptable large increases in the number of false positive cases. In the final analyses, only abnormal heartrate was found to improve triage for dermatological and neurologic/psychiatric presentations.

Respiratory rate showed a significant association with the reference standard in the Ear, Nose and Throat (ENT), gastrointestinal, neurologic/psychiatric, respiratory, trauma or muscular, and general malaise clinical presentations. Optimal cut-offs for the discriminator very abnormal respiratory rate (high urgency) were defined at ≥55 breaths per minute for children <1, and ≥45 breaths per minute for children ≥1 year, and ≥45 breaths per minute and ≥35 breaths per minute for the abnormal respiratory rate discriminator (intermediate urgency). All discriminators improved triage in the respiratory and general malaise clinical presentations. The neurologic or psychiatric clinical presentation improved with the abnormal respiratory rate discriminator only.

Capillary refill time showed a significant association with the reference standard in the ENT, gastrointestinal, neurologic/psychiatric, respiratory, and general malaise presentations. Abnormal capillary refill time, applied as an intermediate urgency discriminator, showed the optimal performance, but did not improve triage in any of the presentations in the final analyses and was therefore not selected as discriminator.

Thus, we determined six novel discriminators (Table 2, S3 File). Two discriminators lead to a very urgent classification: "Very abnormal respiratory rate < 1 year", and "Very abnormal respiratory rate ≥1 year". Four discriminators lead to an intermediate urgency classification: "Abnormal heart rate <1 year", "Abnormal heart rate ≥ 1 year", "Abnormal respiratory rate <1 year", and "Abnormal respiratory rate ≥ 1 year". Adding these discriminators to four different clinical presentations, altered 16 MTS flowcharts.

Performance of the modified MTS

Application of the modified MTS with all new discriminators in our test set reclassified 744 patients (2.5%). The number needed to triage was 41 for one reclassification. Compared with the original MTS, undertriage decreased with 200 patients (0.7%) while an additional 536 patients (1.8%) were overtriaged.

Sensitivity improved from 0.66 (95%CI 0.60-0.72) to 0.71 (95%CI 0.66-0.75) for the high, and from 0.67 (95%CI 0.54-0.76) to 0.70 (95%CI 0.58-0.80) for the high and intermediate urgency patients. Specificity decreased from 0.90 (95%CI 0.86-0.93) to 0.89 (95%CI 0.85-0.92) for the high, and from 0.66 (95%CI 0.52-0.78) to 0.63 (95%CI 0.50-0.75) for the high

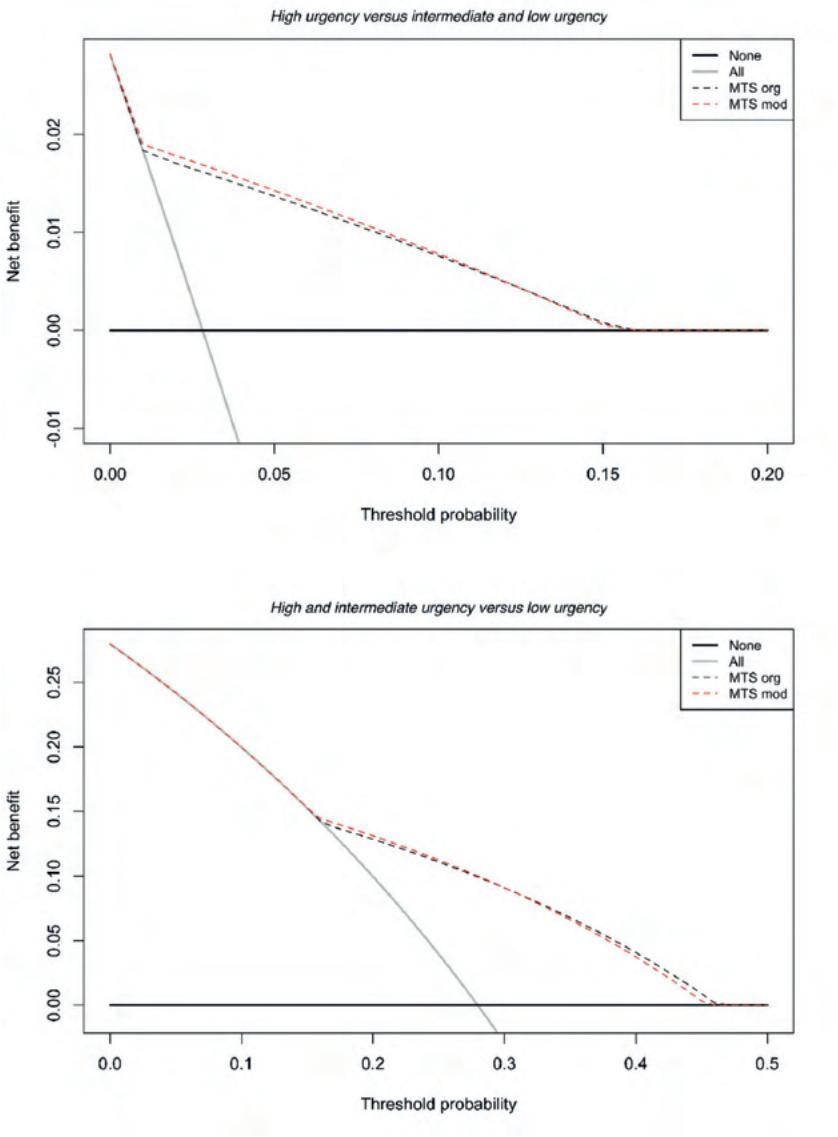
Table 2. Definition and application of new vital sign discriminators

Novel discriminator	Definition	MTS urgency category	Clinical presentation	MTS flowcharts
Very abnormal respiratory rate <1 year	≥ 55 breaths per minute	*Very urgent*	Respiratory General malaise	Asthma, Shortness of breath in children, Unwell child, Irritable child
Very abnormal respiratory rate ≥ 1 year	≥ 45 breaths per minute	*Very urgent*		
Abnormal heart rate <1 year	≥ 160 beats per minute	*Urgent*	Dermatological Neurologic or psychiatric	Rashes, Bites and stings, Burns and scalds, Abscesses and local infections, Wounds, Headache, Fits, Behaving strangely, Overdose and poisoning, Mental illness, Self-harm, Apparently drunk
Abnormal heart rate ≥ 1 year	≥ 140 beats per minute	*Urgent*		
Abnormal respiratory rate <1 year	≥ 45 breaths per minute	*Urgent*	Respiratory General malaise Neurologic or psychiatric	Asthma, Shortness of breath in children, Unwell child, Irritable child, Headache, Fits, Behaving strangely, Overdose and poisoning, Mental illness, Self-harm, Apparently drunk
Abnormal respiratory rate ≥ 1 year	≥ 35 breaths per minute	*Urgent*		

and intermediate urgency patients. Both the increase in sensitivity and decrease in specificity were statistically significant, based on the bootstrapped differences between the original and modified MTS ($p<0.05$,S4 File). According to the decision curves, the modified MTS was the preferred alternative for the range of relevant clinical scenarios (Fig 3). In the ordinal analysis, the modifications improved the R^2 from 0.199 to 0.204.

In the sensitivity analysis in the dataset without imputation, the number of reclassifications based on the vital signs discriminators was 301 (1.0%). The overall improvement remained, although the effect was slightly smaller (S5 File).

Fig 3. Decision curves comparing the modified MTS with the original



DISCUSSION

In this European cohort of more than 100,000 paediatric ED visits, we explored the potential to improve the MTS using vital signs. We developed six novel vital signs discriminators that reduced undertriage when each applied to a specific subgroup of MTS flowcharts. The effect on the triage system as a whole was small, but relevant for the individual patient.

The MTS is a widely implemented triage system in Europe. Previously, in a multicentre study of more than 60,000 children in four European hospitals, we proposed modifications to reduce the proportion of overtriage based on the relocation of existing discriminators and demonstrated their safety.[113] Still, sensitivity of the MTS has been moderate, ranging from 0.65 (95%CI 0.61-0.70) to 0.83 (95%CI 0.79-0.87) in a recent prospective study in three European hospitals.[119] Also several single-centre studies on performance of the MTS in children have shown that undertriage still exist and may have important clinical consequences. .[23, 128]

Physiological measurements have been shown important markers of disease severity in EDs and hospital wards, alone or in combination in early warning scores.[20, 121, 129-132] They are considered crucial in the reliable assessment of any acutely unwell children for the presence of warning signs of underlying serious illness.[102] Moreover, they can be measured routinely by any healthcare professional with experience in the assessment of acutely unwell children. Adding vital signs to a triage system in specific presentations would therefore not greatly affect the ED workflow. Thresholds for abnormal heart rate and respiratory have been proposed[102, 133], but ones that are accurate for and applicable to children presenting with a wide variety of symptoms and optimised for the purpose of triage at the ED remains elusive. A previous study has described the normal ranges and percentiles of vital signs in healthy children.[133] In the ED setting where the majority of children experiences some form of pain, stress or fever, these reference values might be poor predictors of urgency. Clinical reference values such as the APLS guidelines or cut-off values from early warning scores are very heterogeneous and there is no consensus which values should be used in the ED. Moreover, it has been shown that clinical reference values and reference ranges from healthy children are partly overlapping.[133] In a previous large observational study, adding heart rate and respiratory rate to the MTS did not improve its performance.[113] This study, however, added vital signs to all flowcharts in the MTS, and applied pre-defined cut-offs based on previously published 99th percentile values from healthy children.[133] In the current study, we have shown that selecting targeted modifications and using a cut-off value most optimal for the triage setting has the potential to improve the triage of children. The main strength of this study is the large cohort of ED visits from five diverse European EDs. The size of the study population ensured enough power to evaluate modifications in the high urgency population, that only comprises 2% of the study sample. Moreover, we evaluated potential vital signs modifications in a thorough and systematic approach, thereby aiming to identify any relevant discriminators.

However, some limitations have to be acknowledged. In our study vital signs were missing in 42% to 52% of the ED visits. This proportion is in line with previous studies reporting on vital signs measurements in the ED.[103, 134, 135] To deal with the missing values we used a multiple imputation approach to reduce bias.[126] Moreover, a sensitivity analysis in the original dataset without imputation showed that results were largely similar. The number of reclassifications caused by our proposed modifications is relatively small. It is

possible that by limiting the amount of overtriage we considered acceptable, valuable modifications have been missed and a higher sensitivity could have been achieved. Based on the existing structure of the MTS, we selected only two age-specific cut-off values, one for children <1 year and one for children ≥1 year. Although the use more age-categories would have made the cut-off values more precise, we intended to adhere to the MTS' original principles, to facilitate implementation in clinical practice. Paediatric Early Warning Scores (PEWS) are scoring systems based on physiological parameters, that combine the number of abnormal measurements, and the amount of deviation from the normal into a single score.[121, 125] To fully use the potential of vital signs, a PEWS can be used as an additional tool in de triage process, although further research is needed to determine its value in different subgroups of patients. Finally, we used split-sample validation to assess performance of the MTS' modifications and pooled the different performance measures using a random effects model to capture heterogeneity in performance across settings. However, we did not formally externally validate the modified MTS. We propose to evaluate and validate the new modifications after implementation in practice, to gain further understanding of performance in practice in different clinical settings.

The proposed modification results in a decrease of undertriage of 0.7% (n=200 patients) at the cost of an increase in overtriage of 1.8% (n=536). Reducing undertriage is important to improve patient safety. As it decreases delays in care it may prevent morbidity or even mortality. Overtriage does not have a direct impact on patient health and thus, a certain amount of overtriage can be considered acceptable. There is, however, a trade-off. When the amount of overtriage is too large, it will affect waiting times for the truly high urgent patients. In addition, many EDs have limited resources and thus require adequate prioritization. Based on the consensus from the TrIAGE research group, we propose that overtriage of not more than nine patients is acceptable in order to find one additional true high urgent patients, although we acknowledge that in some settings this number may be lower. Our current modification results in the overtriage of less than three patients for each correctly identified high urgent patients, which is well within these pre-specified limits. Since our study used an exhaustive approach we believe it demonstrates the maximal improvement that can be achieved using vital signs within the MTS. The proposed modifications are ready for implementation and validation in clinical practice. Moreover, as we developed targeted modifications to the MTS, the alterations do not have a large impact on the nurses' workload with regard to vital signs' measurements, increasing acceptability in clinical practice. Future studies should focus on novel markers for urgency in children such as nurse or parental gut feeling. Following triage, patients can be monitored for clinical deterioration using re-triage, PEWS[121], or scoring systems for specific subgroups of children[21, 22].

In conclusion, novel age-specific modifications based on the vital signs heart rate and respiratory rate in specific subgroups of flowcharts improve the performance of the MTS. We propose to include these evidence-based modifications in the MTS.

ACKNOWLEDGEMENTS

We gratefully acknowledge the emergency department nursing staff of the participating hospitals for their participation and collection of data. We thank Nienke Seiger, paediatrician (Erasmus MC-Sophia, Rotterdam), Rikke Jorgensen, formerly paediatric research nurse (St. Mary's Hospital, London) and Pinky Rose Espina, paediatric resident (General Hospital, Vienna) for their assistance with data collection for this study. We thank Dorine Borensztajn, paediatrician (Erasmus MC-Sophia, Rotterdam), for her contribution to the development of the TrIAGE database.

SUPPORTING INFORMATION

Available as online web appendix on the website of PLOS One:
- S1 File. Missing vital signs measurements and multiple imputation
- S3 File. Intermediate results step 1 to 3
- S4 File. Performance of modified MTS
- S5 File. Sensitivity analysis based on missing values
- S6 File. TRIPOD checklist

S1 Table. 3-category reference standard as proxy for true patient urgency

3-category reference standard	
High urgency	Mortality at the ED, *and/or* ICU admission immediately after the ED visit, *and/or* Immediate lifesaving interventions*, *and/or* Oxygen administration
Intermediate urgency	Hospital admission immediately after the ED visit, *and/or* IV medication or fluids or inhalation medication at the ED, *and/or* >1 of the following: Radiology; Lab test; Oral medication
Low urgency	None of the above

***** Immediate lifesaving interventions are defined as any of the following :
- airway/breathing support (e.g. intubation or emergent noninvasive positive pressure ventilation);
- electrical therapy (e.g. defibrillation, emergent cardioversion or external pacing);
- emergency procedures (e.g. chest needle decompression, pericardiocentesis, or open thoracotomy)
- haemodynamic support (e.g. significant IV fluid in case of hypotension, blood administration or control of major bleeding) or emergency medications (e.g. atropine, adenosine, inotropics, epinephrine, nalaxon, dextrose in case of hypoglycaemia)

References

Lee JY, Oh SH, Peck EH, et al. The validity of the Canadian Triage and Acuity Scale in predicting resource utilization and the need for immediate life-saving interventions in elderly emergency department patients. Scand J Trauma Resusc Emerg Med 2011;19:68

Platts-Mills TF, Travers D, Biese K, et al. Accuracy of the Emergency Severity Index triage instrument for identifying elder emergency department patients receiving an immediate life-saving intervention. Acad Emerg Med 2010;17(3):238-43.

S2 File. Methodology

The analysis of our study was approached in 4 steps detailed below.

Step 1. Identification of relevant clinical presentations

We grouped all MTS flowcharts into nine clinical presentations: Cardiac, Dermatological, Ear Nose Throat, Gastrointestinal, Neurologic or Psychiatric, Respiratory, Trauma or Muscular, General malaise, Uro- or gynaecological and Other, according to a previous study (Table 1). We choose to combine flowcharts in clinical presentations because this represents the clinical practice of triage where often multiple related flowcharts can be used to adequately triage a patient. Moreover, combining related flowcharts increases the number of patients in the relatively small high urgency category and could improve the power to detect improvement of our modifications.

For each clinical presentation we evaluated an ordinal regression model including MTS urgency level, age (<1 year and ≥1 year), heart rate, respiratory rate and capillary refill time as predictors and the 3-category reference standard as the outcome. Heart rate and respiratory rate were maintained as continuous variables in the analysis. This model assesses the association between vital signs and our outcome measures, adjusted for the already given MTS classification. We applied the likelihood ratio test and considered a p-value of <0.05 statistically significant.

If none of the vital signs in a model was significantly associated with the outcome, it was supposed that this clinical presentation could not be improved by adding vital signs. If one, two or all vital signs were significant with a $p<0.05$ these vital signs were selected for further investigation in relation to the specific clinical presentation.

As a result, three partly overlapping groups were identified: presentations that could potentially be improved with the addition of a heart rate-discriminator, presentations that could potentially be improved with a respiratory rate discriminator and presentations that could potentially be improved with a capillary refill time discriminator.

Table 1. Grouping of flowcharts into clinical presentations

Clinical presentation	MTS Flowcharts
Cardiac	Chest pain, Palpitations
Dermatological	Rashes, Bites and stings, Burns and scalds, Abscesses and local infections, Wounds
Ear, Nose and Throat	Sore throat, Facial problems, Ear problems
Gastrointestinal	Abdominal pain in adults, Abdominal pain in children, Diarrhoea and vomiting, GI bleeding
Neurologic or Psychiatric	Headache, Fits, Collapsed adult, Behaving strangely, Overdose and poisoning, Mental illness, Self-harm, Apparently drunk
Respiratory	Asthma, Shortness of breath in adults, Shortness of breath in children
Trauma or muscular	Limb problems, Major trauma, Neck pain, Back pain, Torso injury, Falls, Assault, Head injury, Limping child
General malaise	Unwell adult, Unwell child, Irritable child
Uro- or gynaecological	Urinary problems, Testicular pain, PV bleeding, Sexually acquired infection, Pregnancy
Other	Major incidents primary, Worried parent, Dental problems, Exposure to chemicals, Foreign body, Diabetes, Eye problems, Allergy, Crying baby, General/Other

Step 2. Defining the optimal cut-off

The vital signs heart rate and respiratory rate were previously maintained continuous but required a cut-off to define which values were considered abnormal. For these vital signs, a cut-off for both high and intermediate urgency was needed. Capillary refill time was available as a dichotomous variable (normal or abnormal). For this vital sign, the analysis should determine whether a positive discriminator should lead to a high and intermediate urgency.

In the analysis, we combined all relevant clinical presentations for each vital sign. Thus, for the analysis of heart rate, we combined all presentations where heart rate had the potential to improve triage according to the analysis in step 1. This was needed because in all different presentations the cut-off value of a vital sign should be the same according to the principles of the MTS.

We analysed the cut-off for heart rate and respiratory rate separately in children <1 year and ≥1 year, thereby aiming to take into account the age-related normal ranges of the vital signs. Moreover, this is consistent with the fever discriminator currently used in the MTS that gives a different priority to children <1 year and ≥1 year.

For a potential modification leading to a high urgency classification, we dichotomised both the MTS and reference standard in a high urgency versus an intermediate and low urgency group. For a potential modification leading to an intermediate urgency classification, we dichotomised the MTS and reference standard in a high and intermediate urgency versus a low urgency group.

We started each analysis with a cross table assessing the association between the dichotomized MTS and the dichotomised reference standard. We calculated the number of total positive patients, true positive patients, and false positive patients. Next, we simulated a modification where a vital sign above a certain cut-off value would place patients in the high urgency level. This process was repeated for the range of relevant cut-off values, with increasing steps of 10 beats per minute for heat rate and 5 breaths per minute for respiratory rate. For each simulated modification we calculated the increase in the total number of positive patients. In addition, we calculated the additional true positive patients and the additional false positive patients and the ratio true positive : false positive patients. We selected the optimal cut-off value according to three principles, based on consensus from the research team. First, we considered only thresholds with a maximum of 20% increase in the total number of positive patients. This, because the number of patients that can be seen immediately in the high urgency categories is limited, and a too large increase in this category makes the triage system inefficient. Second, we limited the ratio additional true positives : additional false positives. For the high urgency discriminators, we found a ratio 1 true positive : 15 false positives the maximum acceptable, for the high to intermediate urgency categories this was 1 true positives : 10 false positives. If multiple cut-off values were appropriate, we selected the cut-off with the largest increase in true positive patients.

Step 3. Selection of the final modifications

In the third step, we simulated that the novel discriminators were added to the MTS, to determine in which clinical presentations the triage was improved. To evaluate performance, in each of the clinical presentation we evaluated an ordinal regression model assessing the association between the original MTS and the reference standard and subsequently between the modified MTS and the reference standard. We performed analyses for each of the new discriminators separately. We selected models with a performance that was better than the original MTS as defined by a higher R^2. We choose the R^2 based on its statistical properties and because it is one of the few performance measures applicable to ordinal models. Ultimately, this selection resulted in a final set of proposed modifications: novel vital-sign discriminators and a subgroup of clinical presentations where triage could be improved.

Step 4. Establishing the performance of the modified MTS

Finally, we assessed the performance of the modified MTS, i.e. the MTS with the new vital signs discriminators, as compared with the original MTS. We applied diagnostic accuracy measures and constructed a decision curve based on a dichotomous analysis, and calculated Nagelkerke's R^2 in an ordinal analysis. These measures were selected based on our previous study evaluating performance measures in the assessment of modifications for triage systems. We constructed 2x2 contingency tables for the original and modified MTS versus the reference standard, and calculated sensitivity, specificity, positive and negative likelihood ratios. For the performance of the high urgency discriminators, we dichotomized the MTS into the high urgency category versus the combination of the intermediate and low urgency categories. For the performance of the intermediate urgency categories, we dichotomized the MTS into the combination of the high and intermediate urgency categories versus the low urgency category. The reference standard was dichotomized in the same manner. The results were calculated for each of the hospitals individually, and pooled using a random effects model. To determine statistical significance, we used bootstrapping to calculate the differences between sensitivity and specificity of the original and modified MTS in a random sample with replacement and repeated this process 1000 times. We calculated the bootstrapped confidence intervals and p-value.

In addition, we constructed decision curves comparing the original and modified MTS. Decision curves provide additional information about clinical value of the proposed modification by incorporating the trade-off between over- and undertriage. In short, the Y-axis represents the benefit of each approach, expressed as net benefit. The highest value indicates the preferred alternative. The X-axis represents the range of preferences one might have regarding not missing any high urgent patients versus the wish to avoid a large increase in overtriage. This is needed because each proposed modification that will correctly assign some ill patients to a high urgency level will also unintentionally

result in some additional low urgency patients falsely triaged to a high urgency level. Although health workers and health care settings may differ as to how much under- and overtriage is acceptable, there is only a range of such values applicable to the clinical practice in emergency departments. We propose that, given the risks involved when missing a high urgency patient, it would be considered unreasonable to accept a more than 20% risk of high urgency to wait before being seen by a physician in any setting. For high and intermediate urgency patients, we propose this risk is at most 40%. Of course, in some settings the acceptable risk might be lower. On the other hand, many emergency departments face large volumes of patients and have only limited capacity to see all high urgency patients at the same time. Therefore, we do not consider overtriage of more than nine patients acceptable in order to find one true high urgency patient. Therefore, we consider the relevant x-axis range between 0.1 and 0.2 for the high urgency patients and between 0.1 and 0.3 for the intermediate urgency patients, and the approach with the highest benefit within this range the preferred alternative.

Finally, we calculated Nagelkerke's R^2 as a marker of overall performance. The calculation was based on an ordinal regression model with MTS urgency (in 3 categories) as the independent variable and the reference standard as the outcome.

References

Zachariasse JM, Seiger N, Rood PP, et al. Validity of the Manchester Triage System in emergency care: A prospective observational study. *PLoS One* 2017;12(2):e0170811

Zachariasse JM, van der Hagen V, Seiger N, et al. Performance of triage systems in emergency care: a systematic review and meta-analysis. *BMJ Open* 2019;9(5):e026471

Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21(1):128-38

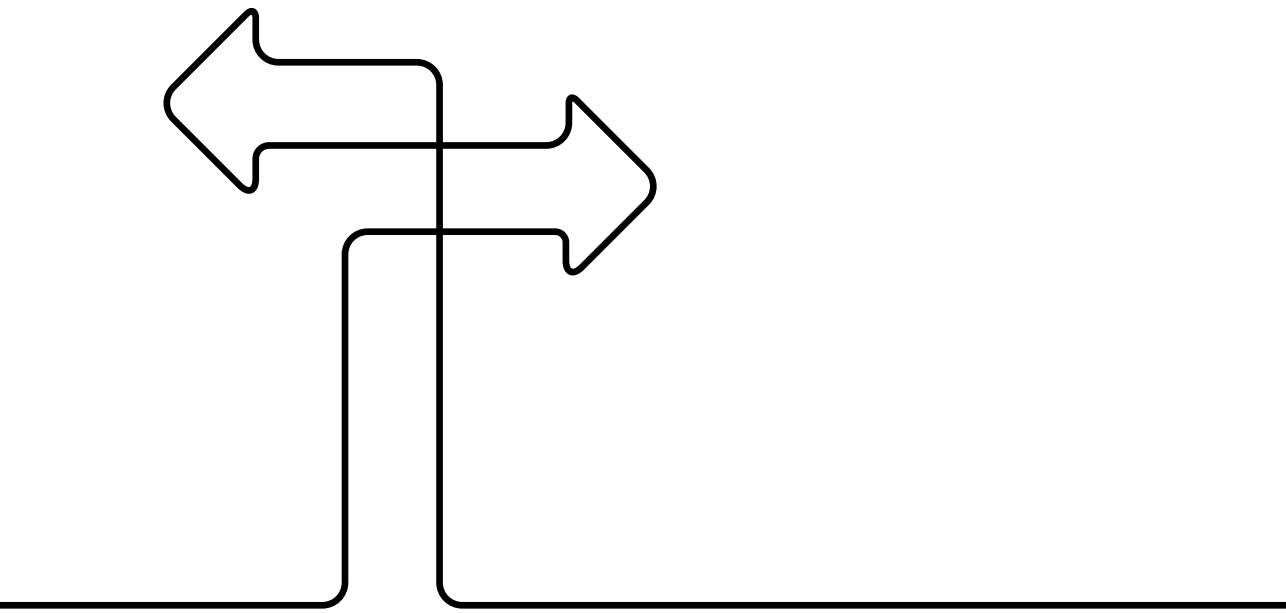
Zachariasse JM, Nieboer D, Oostenbrink R, et al. Multiple performance measures are needed to evaluate triage systems in the emergency department. *J Clin Epidemiol* 2018;94:27-34

Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26(6):565-74

Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 2019;3:18

PART III

New predictors and tools in the initial assessment of children at the emergency department



Chapter 7

The role of clinical impression in the first assessment of children at the emergency department

Joany M Zachariasse, Dominique van der Lee, Nienke Seiger, Evelien de Vos-Kerkhof, Rianne Oostenbrink, Henriëtte A Moll

Arch Dis Child. 2017 Nov;102(11):1052-1056

ABSTRACT

Objective To assess the diagnostic value and determinants of nurses' clinical impression for the recognition of children with a serious illness on presentation to the emergency department (ED).

Design Secondary analysis of a prospective cohort.

Setting and patients 6390 consecutive children <16 years of age presenting to a paediatric ED with a non-surgical chief complaint and complete data available.

Main outcome measures Diagnostic accuracy of nurses' clinical impression for the prediction of serious illness, defined by intensive care unit (ICU) and hospital admission. Determinants of nurses' impression that a child appeared ill.

Results Nurses considered a total of 1279 (20.0%) children appearing ill. Sensitivity of nurses' clinical impression for the recognition of patients requiring ICU admission was 0.70 (95% CI 0.62 to 0.76) and specificity was 0.81 (95% CI 0.80 to 0.82). Sensitivity for hospital admission was 0.48 (95% CI 0.45 to 0.51) and specificity was 0.88 (95% CI 0.87 to 0.88). When adjusted for age, gender, triage urgency and abnormal vital signs, nurses' impression remained significantly associated with ICU (OR 4.54; 95% CI 3.09 to 6.66) and hospital admission (OR 4.00; 95% CI 3.40 to 4.69). Ill appearance was positively associated with triage urgency, fever and abnormal vital signs and negatively with self-referral and presentation outside of office hours.

Conclusion The overall clinical impression of experienced nurses at the ED is on its own, not an accurate predictor of serious illness in children, but provides additional information above some well-established and objective predictors of illness severity.

INTRODUCTION

Early recognition of the critically ill child constitutes a major clinical challenge, particularly at the emergency department (ED). Children visit the ED for a wide variety of illnesses and injuries, ranging in severity from trivial to life-threatening. Signs and symptoms are often non-specific and presentation occurs at different stages of the disease course.[5, 6] Currently, most EDs have triage systems or early warning scores implemented to identify patients with serious conditions at an early stage of the ED visit. Failure to recognise serious illness in children, however, still causes morbidity and mortality. [7, 8]

Nurses' first impression may be a promising additional tool to identify children with serious conditions on presentation to the ED. Clinical impression can be described as the global judgement of a patient's condition, which may be rooted in amongst others knowledge, experience and intuition.[136-138] There is a vast amount of qualitative literature on judgement and decision-making in nursing practice and nursing education. However, quantitative studies assessing the accuracy of health care workers' clinical judgements are scarce.[24, 25]

Because nurses are generally the first point of contact at the ED, use of their overall clinical impression during the initial patient assessment could lead to improved recognition and earlier treatment of the critically ill child. Therefore, the objective of this study is to determine the diagnostic value of nurses' clinical impression at the ED for the recognition of children with a serious illness. Moreover, we aim to identify determinants of this impression to gain a better insight in how nurses judge the condition of their patient, to acknowledge biases in this judgment and to identify new cues for the identification of serious illness in children.

METHODS

This study is based on a secondary analysis of a prospective observational cohort of children presenting to the ED of the Erasmus MC-Sophia Children's Hospital, Rotterdam, the Netherlands.[128, 131] The Medical Ethical Committee of the Erasmus MC approved the study, and the requirement for informed consent was waived (MEC-2016-516).

Study population and data collection

The Erasmus MC-Sophia Children's Hospital is an urban university hospital with approximately 7000 ED visits annually. The ED serves the inner-city population and also holds a regional function for children with significant comorbidity. The ED is staffed by approximately 20 experienced paediatric nurses with an average of more than 10 years of experience in paediatric emergency care. We included all children under the age of 16 years who were seen at the ED by a general paediatrician, between 1 January 2010 and 31 December 2013.

At the ED, nurses routinely recorded elaborate details of each patient visit on structured data entry forms in the electronic medical record. This data entry form has been created for research purposes with the aim to identify determinants and risk factors that can improve decision-making in paediatric emergency care.[21] Standard items include patient demographics, triage classification, presenting symptoms, vital signs and patient disposition.

As part of routine data collection, nurses documented their overall clinical impression of a child during or shortly after triage. Clinical impression was defined as nurses' impression that the child appeared "ill" or "not ill". All study variables were automatically extracted from the patients' electronic health records and entered into an anonymised database.

Data analysis

First, we assessed the diagnostic accuracy of clinical impression for the recognition of serious illness. We defined serious illness as intensive care unit (ICU) admission and as hospital admission immediately after the ED visit. We calculated sensitivity, specificity, likelihood ratios and diagnostic ORs for both outcome measures. To assess the association between nurses' clinical impression and ICU or hospital admission, we performed logistic regression analyses with ICU admission (yes/no) and hospital admission (yes/no) as dependent variables. We included nurses' clinical impression in the model and adjusted the analyses for known predictors of illness severity, including age, gender, triage urgency and presence of abnormal vital signs including heart rate, respiratory rate, oxygen saturation and temperature.[21, 31, 131, 139] Triage is routinely performed in all patients attending the ED using the Manchester Triage System (MTS), the most widely used emergency medical triage system in Europe.[14] The MTS is a flowchart-based algorithm, which classifies patients into one of the five urgency categories. During the triage process, a flowchart is selected based on the patients' reason for encounter (e.g. "shortness of breath" or "abdominal pain"). Flowcharts consist of well-defined signs and symptoms, called discriminators, that determine patients' urgency classification. At the ED of the Erasmus MC-Sophia Children's Hospital, a computerised version of the official Dutch translation of the MTS with validated modifications for children with fever is used.[35, 37] Heart rate and respiratory rate were considered abnormal if they were lower than the 1st or higher than the 99th percentile values published by Fleming *et al*.[133] Abnormal saturation was defined as a peripheral oxygen saturation ≤94% and ≤90%, measured by pulse oximetry, as used in the pediatric early warning score by Parshuram *et al*.[140] Fever was defined as a body temperature of 38°C or higher.[21]

To study determinants of nurses' clinical impression, we performed a multivariable logistic regression analysis with nurses' clinical impression (ill appearing or not ill appearing) as the dependent variable. As determinants, we analysed several clinical factors: age, gender, triage urgency, abnormal heart rate, respiratory rate, oxygen

saturation and temperature. In addition, we evaluated two non-clinical factors: type of referral (self, general practitioner or other) and shift (office hours or out-of-hours).

Missing values

Nurses' clinical impression was available in 58% of ED visits (Table 1). We assumed that these visits comprised a random sample of all ED attendances, because nurses mainly did not complete the data entry forms when they did not have enough time. This is supported by the fact that we did not observe differences in the proportion of mortality in the ED, ICU and hospital admission between the included and excluded children (Pearson's χ^2 (3) =5.75, p=0.13). Therefore, we excluded the cases with missing clinical impression and performed a complete case analysis. Missing vital signs were assumed to be in the normal range of the cut-offs we defined before, which is in line with clinical experience.

In addition, we performed a sensitivity analysis where we repeated our analyses on a dataset in which the missing values were imputed. We created a multiple imputation model including all variables used in the multivariable regression analyses, the outcome variables and several other relevant variables describing case mix of the patients. The imputation process resulted in ten databases, and we calculated the pooled estimates.[141]

Multiple imputation was performed by using the Hmisc package in R.[115] SPSS V.20.0 and the VassarStats website (www.vassarstats.net) were used for the further statistical analysis.

RESULTS

During the study period, 11,024 children visited the ED with a medical illness. Nurses recorded their clinical impression in 6390 children (58%) and 1279 (20%) of these were considered "ill". Of the 6390 included patients, 1170 (18.3%) were admitted to the hospital and 171 (2.7%) were admitted to ICU (Table 1). The clinical characteristics and outcome of the total population were comparable to the included study population. Eight patients died during the ED visit and were excluded from further analysis. All eight patients were triaged to the highest urgency categories (seven "emergent" and one "very urgent"), indicating that they were probably already being resuscitated on arrival and did not need a nursing assessment to establish urgency.

Table 1. Characteristics of the study population

	Total population (n = 11,024)	**Clinical impression documented (n = 6390)**
Age, n (%)		
<1 year	3029 (27.5)	1909 (29.9)
1-2 years	1694 (15.4)	1035 (16.2)
2-5 years	2568 (23.3)	1464 (22.9)
5-12 years	2602 (23.6)	1382 (21.6)
>12 years	1131 (10.3)	600 (9.4)
Gender, n (%)		
Female	4810 (43.6)	2808 (43.9)
Male	6214 (56.4)	3582 (56.1)
Triage urgency, n (%)		
Immediate / Very urgent	2122 (19.2)	1183 (18.5)
Urgent	4757 (43.2)	2745 (43.0)
Standard / Non urgent	3909 (35.5)	2387 (37.4)
Missing	236 (2.1)	75 (1.2)
Referral, n (%)		
Self	4103 (37.2)	2518 (39.4)
General practitioner	2249 (20.4)	1349 (21.1)
Medical specialist	2604 (23.6)	1376 (21.5)
Other	2068 (18.8)	1147 (17.9)
Clinical impression, n (%)		
Ill	1279 (11.6)	1279 (20.0)
Not ill	5111 (46.4)	5111 (80.0)
Missing	4634 (42.0)	N/A
Disposition, n (%)		
Death in the ED	17 (0.2)	8 (0.1)
ICU admission	325 (2.9)	171 (2.7)
Hospital admission	2046 (18.6)	1170 (18.3)
Discharge / other	8636 (78.3)	5041 (78.9)

Sensitivity of nurses' clinical impression for the recognition of patients requiring ICU admission was 0.70 (95% CI 0.62 to 0.76) with a negative likelihood ratio of 0.37 (95% CI 0.30 to 0.47). Sensitivity for hospital admission was 0.48 (95% CI 0.45 to 0.51) with a negative likelihood ratio of 0.59 (95% CI 0.56 to 0.63). This corresponds to a low rule-out value for the presence of serious illness.

The rule-in value for the presence of serious illness was moderate, with a specificity for ICU admission of 0.81 (95% CI 0.80 to 0.82) and a positive likelihood ratio of 3.74 (95% CI 3.35 to 4.19). Specificity for hospital admission was 0.88 (95% CI 0.87 to 0.88) with a positive likelihood ratio of 3.84 (95% CI 3.51 to 4.21) (Table 2).

Table 2. Diagnostic value of nurses' clinical impression for the prediction of ICU admission or hospital admission (n=6382)

	ICU admission	Hospital admission
Sensitivity (95% CI)	0.70 (0.62 to 0.76)	0.48 (0.45 to 0.51)
Specificity (95% CI)	0.81 (0.80 to 0.82)	0.88 (0.87 to 0.88)
Positive Likelihood Ratio (95% CI)	3.74 (3.35 to 4.19)	3.84 (3.51 to 4.21)
Negative Likelihood Ratio (95% CI)	0.37 (0.30 to 0.47)	0.59 (0.56 to 0.63)
OR, univariable (95% CI)	10.02 (7.19 to 13.96)	6.47 (5.65 to 7.41)
OR, multivariable (95% CI)[a,b]	4.54 (3.09 to 6.66)	4.00 (3.40 to 4.69)

[a] Patients with missing data for any of the predictor variables (n=75) are excluded
[b] Associations determined by logistic regression with no ICU or no hospital admission as the reference group. The multivariable model is adjusted for age, gender, triage urgency, abnormal respiratory rate, heart rate or oxygen saturation and fever.

When adjusted for age, gender, triage urgency, fever or abnormal vital signs, nurses' impression that the child appeared ill remained significantly associated with ICU admission (OR 4.54; 95% CI 3.09 to 6.66) and hospital admission (OR 4.00; 95% CI 3.40 to 4.69) (Table 2).

About 2387 patients were triaged as low urgent according to MTS categories 4 and 5. In this group, 11 patients were admitted to ICU and 247 to hospital. In this subset of patients, nurses' clinical impression was important for signalling, but not for excluding serious illness. The positive likelihood ratio for the detection of ICU admission increased from 3.74 (95% CI 3.35 to 4.19) to 4.15 (95% CI 1.17 to 14.76) and remained similar for the detection of hospital admission. Nurses' clinical impression was able to identify 2 (18%) additional ICU admissions and 32 (13%) additional hospital admissions in patients with a low triage urgency. Yet, this would come at the cost of a poor negative likelihood ratio and a high absolute number of false alarms, if we would use ill appearance as the only marker (Table 3).

Table 3. Nurses' clinical impression in children triaged as low urgent (n=2387)

	ICU admission	No ICU admission	Total
Child ill appearing	2	104	106
Child not ill appearing	9	2272	2281
Total	11	2376	2387
	Hospital admission	**No hospital admission**	**Total**
Child ill appearing	32	74	106
Child not ill appearing	215	2066	2281
Total	247	2140	2387

ICU, intensive care unit

In addition, we evaluated determinants of nurses' clinical impression that a child appeared ill. We found a positive association with high triage urgency, and the presence of abnormal vital signs, and a negative association with self-referral and presentation out-of-hours (Table 4).

Table 4. Determinants of nurses' clinical impression

	n (%)[a]	OR, univariable (95% CI)[b]	OR, multivariable (95% CI)[b]
Age			
<1 year	1879 (29.8)	1.03 (0.80 to 1.31)	1.06 (0.80 to 1.40)
1-2 years	1025 (16.3)	1.64 (1.27 to 2.11)	1.12 (0.84 to 1.50)
2-5 years	1449 (23.0)	1.38 (1.08 to 1.76)	1.09 (0.83 to 1.45)
5-12 years	1363 (21.6)	1.06 (0.82 to 1.37)	1.00 (0.75 to 1.34)
>12 years	591 (9.4)	*Reference*	*Reference*
Gender			
Female	2769 (43.9)	0.81 (0.71 to 0.92)*	0.87 (0.75 to 1.00)
Male	3538 (56.1)	*Reference*	*Reference*
Triage urgency			
Immediate / Very urgent	1175(18.6)	19.40 (15.47 to 24.31)*	9.52 (7.48 to 12.12)*
Urgent	2745 (43.5)	6.03 (4.87 to 7.48)*	4.13 (3.30 to 5.16)*
Standard / Non urgent	2387 (37.8)	*Reference*	*Reference*
Referral			
Self	2492 (39.5)	0.41 (0.34 to 0.49)*	0.57 (0.46 to 0.69)*
General practitioner	1334 (21.2)	0.80 (0.67 to 0.96)*	0.94 (0.76 to 1.16)
Medical specialist	1352 (21.4)	0.63 (0.52 to 0.75)*	0.87 (0.70 to 1.08)
Other	1129 (17.9)	*Reference*	*Reference*
Respiratory Rate			
Normal	5538 (87.8)	*Reference*	*Reference*
Abnormal	769 (12.2)	2.78 (2.37 to 3.27)*	1.46 (1.20 to 1.78)*
Oxygen saturation			
Normal	6001 (95.1)	*Reference*	*Reference*
Low (91% - ≤94%)	173 (2.7)	5.28 (3.89 to 7.17)*	2.32 (1.63 to 3.30)*
Very low (≤90%)	133 (2.1)	8.88 (6.16 to 12.79)*	3.54 (2.35 to 5.33)*
Heartrate			
Normal	5229 (82.9)	*Reference*	*Reference*
Abnormal	1078 (17.1)	5.11 (4.43 to 5.89)*	2.05 (1.73 to 2.43)*
Fever			
Absent	4547 (72.1)	*Reference*	Reference
Present	1760 (27.9)	4.14 (3.64 to 4.72)*	2.52 (2.16 to 2.95)*
Shift			
Office hours	2809 (44.5)	*Reference*	*Reference*
Out-of-hours	3498 (55.5)	1.08 (0.95 to 1.22)	0.75 (0.65 to 0.87)*

[a] Patients with missing data for any of the predictor variables (n=75) are excluded
[b] Associations determined by logistic regression with patients appearing "not ill" as the reference group. The multivariable model includes all predictors
* Indicates predictors where the CI does not include 1

Using the multiple imputation approach, our results were comparable to those from the primary analysis and our conclusions remained robust (see online supplemental table 1 and 2). Therefore, we believe the risk of bias due to the missing data in our study is limited.

DISCUSSION

The clinical impression of experienced nurses during their first assessment at the ED is by itself not an accurate predictor of serious illness in children. Although the rule-in value is moderate, the poor rule-out value indicates that seriously ill children may be missed. And while additional seriously ill children may be identified, this will come at the cost of a considerable number of false alarms. Therefore, nurses' clinical impression should not be used on its own to determine the severity of illness of a child. However, when nurses' clinical impression is added to a model with age, gender, triage urgency and abnormal vital signs, it remains significantly associated with serious illness. This indicates that nurses' first impression has incremental value above other commonly used and objective predictors of illness severity at the ED.

To gain more insight in the factors that contribute to nurses' first overall impression, we evaluated the association with several clinical and non-clinical variables. Our results demonstrate that nurses' clinical impression at the ED is positively associated with triage urgency and abnormal vital signs. As expected, these commonly used and objective markers contribute to nurses' impression that a child appears ill. Surprisingly, there is no association between clinical impression and age, indicating that nurses base their judgment on signs and symptoms regardless of the age of the child. A negative association exists between nurses' clinical impression and self-referral and presentation out-of-office hours. This could indicate that nurses correctly take these determinants into account as markers of lower illness severity, but it could also represent a bias in nurses' clinical judgment.

In the past decades, multiple qualitative studies have described and explored the concept of decision-making in nursing.[24, 25] However, quantitative research on the diagnostic accuracy of healthcare workers' clinical impression in emergency care is scarce. A study by Van den Bruel *et al* found an association between presence of primary care physicians' gut feeling and serious bacterial infections in children.[139, 142] Another study found an association between nurses' clinical impression and severity of sepsis of ED adult patients.[143] To the best of our knowledge, the current study is the first to evaluate nurses' clinical impression in emergency care for the recognition of seriously ill children.

A limitation of our study is that it consists of a secondary analysis of prospectively collected data. Although we attempted to maximize the registration of routine data during the study period, clinical impression was not documented in 42% of patients. Missing data may lead to biased results and therefore we used two different approaches to analyse the data: an approach where we excluded patients with missing clinical

impression and a multiple imputation approach. Both gave comparable results and therefore, we believe a true association between nurses' clinical impression and serious illness in children exists.

Because the current study was not planned in advance, we did not obtain data on nurse characteristics, such as age or years of experience. However, most nurses at our ED are highly specialised with numerous years of working experience in paediatric emergency care. Hence, we would not have been able to compare nurses with different levels of experience. Furthermore, because the study was performed in one university hospital - tertiary referral centre, generalisability of our findings to general or teaching hospitals is unknown.

Preferably, this study should be repeated prospectively and in different settings to confirm the results and assess its generalisability. Currently, nurses' overall clinical impression was not included as a separate item in most paediatric early warning scores or ED triage systems;[14, 131] although some include more general items such as "worried about clinical state"[144] or "high risk situation"[11] or recommend using the Paediatric Assessment Triangle[10]. While our results are preliminary, further extension of this work could evaluate the incremental value of overall impression in, for example, early warning scores, triage systems or prediction models. In addition, research should aim at identifying the determinants of nurses' clinical impression, including patient and nurse characteristics. Understanding how nurses' form their first impression can lead to new predictors of illness severity and improved recognition of critically ill children. Finally, it would be worthwhile to explore the role of parents' impression of their child, to improve the recognition of serious illness in the ED.

CONCLUSION

This study shows that the first clinical impression of experienced nurses is significantly associated with serious illness in children as reflected by ICU and hospital admission, even when age, gender, triage urgency, and abnormal vital signs are taken into account. Although in itself not an accurate predictor, nurses' clinical impression appears to provide additional information above some well-established predictors of illness severity. These results suggest that nurses' overall impression could be an interesting additional tool in the first assessment of children at the ED.

ACKNOWLEDGEMENTS

We gratefully acknowledge the nurses of our emergency department for their participation and careful collection of the data for this study.

SUPPORTING INFORMATION

Table S1. Diagnostic value of nurses' clinical impression for the prediction of ICU admission or hospital admission, using the multiple-imputation approach (n=11,007)

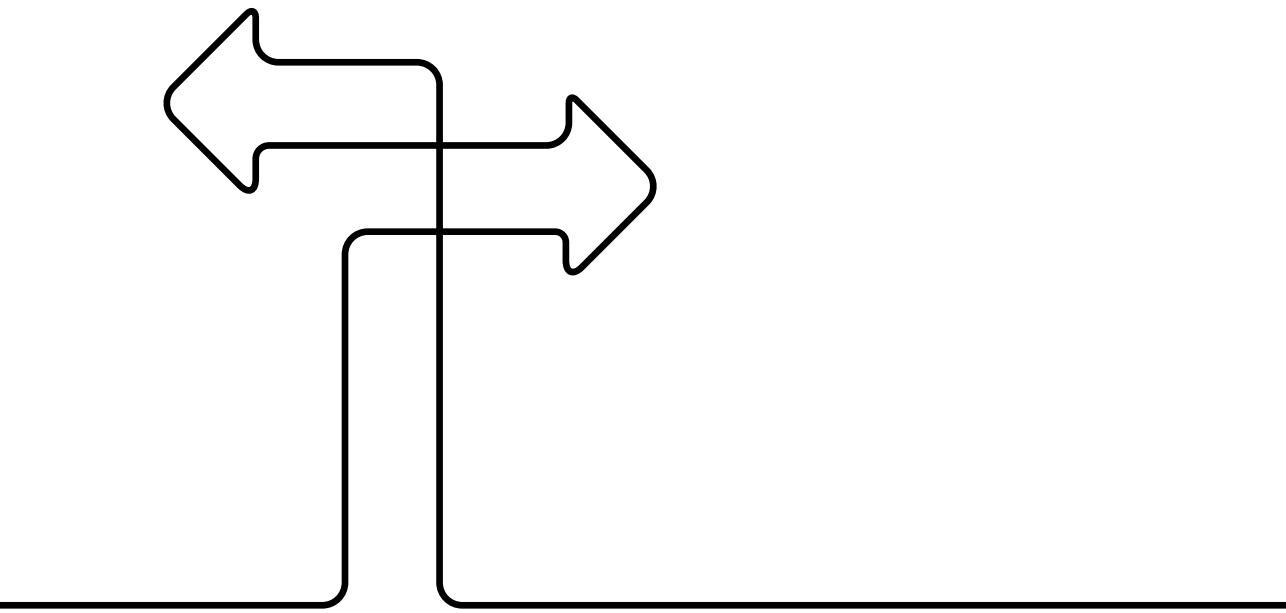
	ICU admission	Hospital admission
Sensitivity (95% CI)	0.64 (0.57-0.71)	0.48 (0.46-0.51)
Specificity (95% CI)	0.81 (0.77-0.85)	0.87 (0.86-0.89)
Positive Likelihood Ratio (95% CI)	3.38 (2.92-3.91)	3.83 (3.48-4.21)
Negative Likelihood Ratio (95% CI)	0.44 (0.35-0.57)	0.59 (0.57-0.62)
Odds Ratio, Univariable (95% CI)	7.61 (5.18-11.18)	6.46 (3.24-12.89)
Odds ratio, Multivariable (95% CI)[a]	3.76 (2.33-6.06)	4.16 (3.43-5.05)

[a] Associations determined by logistic regression with no ICU or no hospital admission as the reference group. The multivariable model is adjusted for age, gender, triage urgency, abnormal respiratory rate, heart rate or oxygen saturation and fever.

Table S2. Determinants of nurses' clinical impression, using the multiple-imputation approach (n=11,007)

	OR, univariable (95% CI)[a]	OR, multivariable (95% CI)[a]
Age		
<1 year	0.97 (0.78-1.20)	1.02 (0.78-1.32)
1-2 years	1.37 (1.08-1.74)*	1.01 (0.76-1.35)
2-5 years	1.15 (0.93-1.43)	0.94 (0.72-1.22)
5-12 years	0.99 (0.79-1.24)	0.95 (0.73-1.25)
>12 years	*Reference*	*Reference*
Gender		
Female	0.85 (0.76-0.95)*	0.90 (0.79-1.02)
Male	*Reference*	*Reference*
Triage urgency		
Immediate / Very urgent	15.66 (12.78-19.18)*	9.49 (7.69-11.70)*
Urgent	4.69 (3.88-5.67)*	3.53 (2.89-4.32)*
Standard / Non urgent	*Reference*	*Reference*
Referral		
Self	0.42 (0.36-0.50)*	0.61 (0.51-0.72)*
General Practitioner	0.78 (0.66-0.94)*	0.97 (0.79-1.19)
Medical specialist	0.60 (0.50-0.70)*	0.84 (0.70-1.01)
Emergency service / Other	*Reference*	*Reference*
Respiratory Rate		
Normal	*Reference*	*Reference*
Abnormal	1.69 (1.48-1.93)*	1.17 (0.98-1.41)
Oxygen saturation		
Normal	*Reference*	*Reference*
Low (≤94%)	3.50 (2.64-4.65)*	2.13 (1.53-2.98)*
Very low (≤90%)	5.16 (3.89-6.86)*	3.00 (2.18-4.14)*
Heartrate		
Normal	*Reference*	*Reference*
Abnormal	3.44 (2.97-3.99)*	1.71 (1.45-2.01)*
Fever		
Absent	*Reference*	*Reference*
Present	3.18 (2.79-3.62)*	2.12 (1.81-2.48)*
Shift		
Office hours	*Reference*	*Reference*
Out-of-hours	0.94 (0.85-1.05)	0.76 (0.67-0.86)*

[a] Associations determined by logistic regression with patients appearing "not ill" as the reference group. The multivariable model is adjusted for age, gender, triage urgency, referral type, abnormal respiratory rate, heart rate or oxygen saturation, fever, and nursing shift.
* Indicates predictors where the CI does not include 1



Chapter 8

Association between hypotension and serious illness in the emergency department: an observational study

Nienke N Hagedoorn, Joany M Zachariasse, Henriëtte A Moll

Arch Dis Child. 2020 Jun;105(6):545-551

ABSTRACT

Background The value of routine blood pressure measurement in the emergency department (ED) is unclear.

Objective To determine the association between hypotension in addition to tachycardia and the Shock Index for serious illness.

Design Observational study.

Setting University ED (2009-2016).

Participants, methods and main outcomes Routine data collected from consecutive children <16 years. Using logistic regression, we assessed the association between hypotension (adjusted for tachycardia) and Shock Index (ratio heart rate/blood pressure [BP]) for serious illness. The predictive accuracy (sensitivity, specificity) for hypotension and Shock Index was determined for serious illness, defined as intensive care unit (ICU) and hospital admissions.

Results We included 10 698 children with measured BP. According to three age-adjusted clinical cut-offs (Advanced Paediatric Life Support, Paediatric Advanced Life Support and Paediatric Early Warning Score), hypotension was significantly associated with ICU admission when adjusted for tachycardia (range OR 2.6-5.3). Hypotension showed low sensitivity (range 0.05-0.12) and high specificity (range 0.95-0.99) for ICU admission. Combining hypotension and tachycardia did not change the predictive value for ICU admission. Similar results were found for hospitalisation. Shock index was associated with serious illness. However, no specific cut-off value was identified in different age groups.

Conclusions Hypotension, adjusted for tachycardia, is associated with serious illness, although its sensitivity is limited. Shock index showed an association with serious illness, but no acceptable cut-off value could be identified. Routine BP measurement in all children to detect hypotension has limited value in the ED. Future studies need to confirm which patients could benefit from BP measurement.

INTRODUCTION

Vital signs are essential for recognising serious illness in children in the emergency department (ED). However, the frequency of blood pressure (BP) measurement varies widely (23%-87%) and no consensus exists on performing routine BP measurement to detect hypotension.[26, 27, 145] Accurate age-related cut-offs are needed to assess hypotension as incorrect cut-offs may lead to false-positive or false-negative results. Although paediatric guidelines provide different definitions of low BP, it is unclear which BP cut-off should be used in the ED.[28, 146, 147]

Moreover, the predictive value of hypotension for serious illness is unclear in the diverse ED population. In children, hypotension is considered a late sign of deterioration and is used for diagnosis of shock. Children increase heart rate to preserve cardiac output.[148, 149] Since abnormal heart rate occurs in an earlier phase, the additional value of routine BP over heart rate in prediction of serious illness could be limited in the ED.

Another measure of haemodynamic status is Shock Index, the ratio of heart rate to systolic BP, which is associated with mortality and disease severity in adults.[150-152] In small cohorts of children, elevated Shock Index has been associated with injury severity in trauma and mortality in septic shock.[153-156] However, the shock index in all paediatric ED patients has not yet been evaluated and could be an important predictor in children.

This study aims to study the additional value of BP measurement: 1) To determine the predictive capability of hypotension in addition to tachycardia. 2) To assess the utility of Shock Index for serious illness in children. This observational study is based on routine BP measurements in the ED using electronic health records.

METHODS

Design

We applied three commonly used clinical definitions for hypotension on data from a prospective study of children visiting the ED to determine the predictive value of hypotension in addition to tachycardia for serious illness. Second, we studied the predictive ability of the Shock Index. This was a secondary analysis in a study validating the Manchester Triage System (MTS).[119, 128]

Setting

The observational study included all children <16 years who presented consecutively at the ED of Erasmus MC-Sophia Children's Hospital (Rotterdam, The Netherlands) between August 2009 and December 2016. This inner-city university hospital receives approximately 7,000 children annually.

Data collection

Data of patient characteristics, vital signs, triage level and disposition were automatically derived from electronic health records that were completed by trained nurses during triage. Heart rate was measured using pulse oximeters and BP using the oscillometric infinity M540 monitor (Draeger Medical, Telford, Pennsylvania, USA). BP was measured on medical indication at the discretion of the nurse or attending physician.

Outcomes and definitions

Serious illness was defined as admission to the ICU or hospital following ED visit. Indications for ICU admission include requirement of advanced respiratory support ([non-] invasive ventilation, high flow oxygen); inotropes or continuous intravenous antiepileptics; tracheal cannula; acute or threatening failure of more than two organ systems which was expected to last >24 hours or in a child <1 year.[128] We selected three age-adjusted clinical cut-offs to define hypotension to demonstrate the range in clinical practice: Advanced Paediatric Life Support (APLS)[157], Paediatric Advanced Life Support (PALS)/septic shock screening tool[158, 159], and the Paediatric Early Warning score (PEWS) [160](table 1). Heart rate was categorized as tachycardia versus no tachycardia according to the same reference as the BP cut-off (online supplementary appendix 1). Children with bradycardia (5.9%-7.4%) were defined as no tachycardia. Age was categorised as 0-1 year, 1-2 years, 2-5 years, 5-12 years and 12-16 years. Triage urgency was determined by MTS V.3.[35] Ill appearance was assessed by the nurse on a 2-point scale: ill versus non-ill appearance.

Data analysis

Our sample was limited to patients with measured heart rate and BP. Children who died in the ED were excluded (n=34). The value of BP measurement could be limited in this group, since the majority (94%) was triaged as emergencies. Outliers were verified in patient records. First, we assessed the relation between BP and heart rate, using scatterplots. To facilitate analysis across age-groups, we standardised heart rate and BP using z-scores, which were calculated separately for the different age categories. Second, we assessed the association between hypotension and serious illness using the three clinical cut-offs for hypotension. We used univariable logistic regression to evaluate the association of different BP cut-offs with ICU or with hospital admission, and adjusted for tachycardia in a multivariable model.

We determined the predictive value of hypotension for ICU admission and hospitalisation by calculating sensitivity, specificity, positive and negative likelihood ratios.[161] To study the predictive value of hypotension in addition to tachycardia, we calculated the predictive value of 1) hypotension; 2) tachycardia; 3) the combination of tachycardia and hypotension; and 4) either hypotension or tachycardia. Positive likelihood ratios >5 and negative likelihood ratios <0.2 were considered relevant.[162]

Table 1. Definition of hypotension per different age groups for systolic blood pressure in mm Hg

Age range	APLS	PEWS	PALS
<4 weeks	<75	≤60	<60
4 weeks - 6 weeks	<75	≤60	<70
6 weeks - 3 months	<75	≤60	<70
3-6 months	<75	≤80	<70
6-12 months	<75	≤80	<70
1-2 years	<75	≤90	<72
2-3 years	<80	≤90	<74
3-4 years	<80	≤90	<76
4-5 years	<80	≤90	<78
5-6 years	<90	≤90	<80
6-7 years	<90	≤90	<82
7-8 years	<90	≤90	<84
8-9 years	<90	≤90	<86
9-10 years	<90	≤90	<88
10-12 years	<90	≤90	<90
12-13 years	<105	≤100	<90
13-14 years	<105	≤100	<90
14-16 years	<105	≤100	<90

APLS, advanced paediatric life support; PEWS, pediatric early warning score; PALS, pediatric advanced life support

The normal range of Shock Index (ratio of heart rate to BP) is age dependent.[163] Therefore, we stratified the analysis for Shock Index by age. To assess the association of Shock Index, we used univariable logistic regression. To facilitate interpretation, the OR present the odds for 0.1 unit increase in shock index. Next, the discriminative ability was presented by the area under the curve (AUC) of receiver operating characteristics. We used Youden's index to identify the optimal cut-off value to assess the predictive value. [164] We merged the age groups into <2 years, 2-10 years and >10 years to ensure sufficient numbers for statistical analysis. To explore age-adjusted cut-off values for high Shock Index, we defined a cut-off by dividing the APLS tachycardia value with the APLS hypotension value for each age group (online supplementary appendix 2).

Subgroup analyses were performed in patients with ill appearance, fever (temperature >38.0°C) and patients presenting with surgical problems including major trauma, head injury, limb problems, wounds, torso injuries and assault [31].

Data analyses and visualisation were performed in SPSS V.24.0 and R. The medical ethical committee waived the requirement for informed consent.

RESULTS

During the study period, 45,495 children (58.6% male) presented to the ED; 891 (2.0%) were triaged as emergencies. A total of 10,698 patients had BP and heart rate measured. In this sample, 3907 (36.5%) children were admitted to the general ward and 631 (5.9%) were admitted to the ICU (table 2). Patients with BP measurement were older, had higher urgency level and were more often admitted compared with children without BP measurement (online supplementary appendix 3). The prevalence of hypotension ranged from 1.2% to 5.3% depending on the cut-off used (online supplementary appendix 4). In children with hypotension according to APLS, 13.9% were admitted to the ICU and 33.5% were hospitalized.

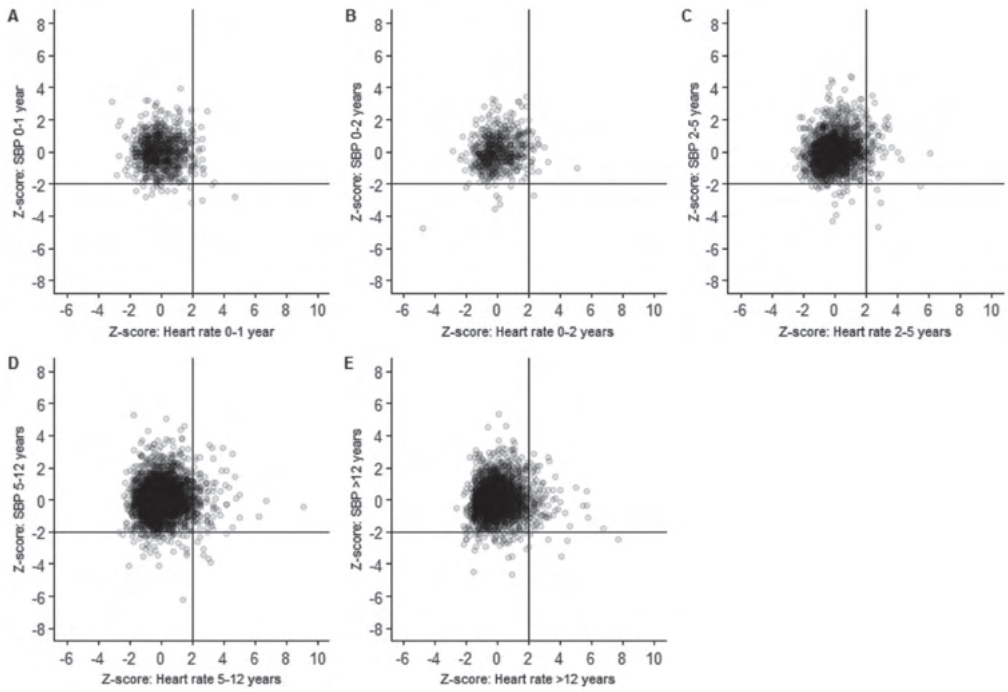
Table 2. Characteristics of visits at the paediatric emergency department of Sophia Children's Hospital from 2009 to 2016

	Total	**Patients with blood pressure and heart rate measured**	**Patients with hypotension according to APLS**
	n=45,495	**n=10,698**	**N=504**
Male; n %	26338 (57.9)	5872 (54.9)	219 (43.5)
Age in years; median, (IQR)	4.3 (1.4 - 9.8)	7.74 (3.6 – 7.7)	13.0 (6.67 – 14.5)
Age category; n (%)			
0 - 1 year	8734 (19.2)	920 (8.6)	78 (15.5)
1 - 2 years	5736 (12.6)	668 (6.2)	7 (1.4)
2 - 5 years	10154 (22.3)	2091 (19.5)	14 (2.8)
5 - 12 years	13503 (29.7)	4101 (38.3)	80 (15.9)
12 - 16 years	7368 (16.2)	2918 (27.3)	325 (64.5)
MTS urgency; n (%)			
Emergent / Very urgent	6433 (14.2)	2572 (24.0)	155 (30.7)
Urgent	19873 (43.7)	5026 (47.0)	199 (39.5)
Standard/ Non urgent	17711 (38.9)	2922 (27.3)	163 (27.0)
Missing	1478 (3.2)	178 (1.7)	14 (2.8)
Disposition; n (%)			
Admission general ward	8848 (19.4)	3276 (30.6)	169 (33.5)
Intensive care	1132 (2.5)	631 (5.9)	70 (13.9)
Died	34 (0.1)	- *	- *
Discharge	34913 (76.7)	6719 (62.8)	261 (51.8)
Other	401 (0.9)	61 (0.6)	4 (0.8)
Missing	167 (0.4)	11 (0.1)	0 (0.0)
Shock index; mean (sd)			
0 - 1 year		1.52 (0.48)	
1 - 2 years		1.25 (0.31)	
2 - 5 years		1.11 (0.26)	
5 - 12 years		0.89 (0.24)	
12 - 16 years		0.76 (0.22)	

* Children who died were excluded
APLS, advanced paediatric life support; MTS, Manchester Triage System

Our study found no association between z-scores of heart rate and BP in any of the age categories (Pearson correlation 0.04-0.18) (figure 1). In particular, no clear relation was observed between low BP and high z-scores for heart rate.

Figure 1. Scatter plots of z-scores of heart rate and systolic blood pressure (SBP) for different age categories (A; 0-1 year, B; 1-2 years, C; 2-5 years, D; 5-12 years, E; 12-16 years)



Hypotension, as a sole predictor, had an association with ICU admission (range OR 2.56-5.27) and hospital admission (range OR 1.46–2.66). The association between hypotension and serious illness remained significant after adjustment for tachycardia. In this analysis, the PALS cut-off for hypotension showed the strongest association with ICU admission and hospitalisation (table 3).

Table 3. Logistic regression analysis for ICU and hospital admission

	Patients with hypotension/ tachycardia	ICU admission		Hospital admission	
		OR	95% CI	OR	95% CI
APLS					
Hypotension	n=504	2.77	2.12-3.62	1.61	1.34-1.92
Tachycardia (APLS)	n=1692	2.46	2.06-2.94	2.62	2.36-2.91
Hypotension adjusted for tachycardia		2.68	2.05-3.51	1.56	1.30-1.88
PALS/septic shock screening tool					
Hypotension	n=133	5.27	3.51-7.91	2.66	1.87-3.77
Tachycardia (septic shock screening tool)	n=1709	1.80	1.49-2.18	1.91	1.72-2.12
Hypotension adjusted for tachycardia		4.99	3.32-7.52	2.52	1.77-3.59
PEWS					
Hypotension	n=571	2.56	1.98-3.31	1.46	1.24-1.73
Tachycardia (PEWS)	n=4113	2.02	1.72-2.37	2.16	1.99-2.34
Hypotension adjusted for tachycardia		2.54	1.96-3.29	1.46	1.23-1.73
Shock Index*					
Age 0-1 year		1.09	1.06-1.14	1.14	1.09-1.18
Age 1-2 years		1.07	0.99-1.16	1.07	1.02-1.22
Age 2-5 years		1.08	1.02-1.15	1.06	1.02-1.09
Age 5-12 years		1.13	1.08-1.19	1.14	1.11-1.18
Age >12 years		1.22	1.15-1.29	1.19	1.15-1.24

*ORs present each 0.1 increase in Shock Index
APLS, advanced paediatric life support; ICU, intensive care unit; PALS, Paediatric Advanced Life Support; PEWS, Paediatric Early Warning Score

The cut-offs for hypotension showed a low sensitivity and a high specificity for serious illness (table 4). For ICU admission, specificity ranged between 0.95 and 0.99 and sensitivity between 0.05 and 0.12. The positive likelihood ratios ranged from 2.38 to 5.06 and the negative likelihood ratios ranged from 0.93 to 0.96. The combination of tachycardia and hypotension did not improve the performance for ICU admission with low sensitivity (0.02-0.08) and high specificity (0.94-0.98). The analysis for hospital admission showed similar results.

Average values for Shock Index decreased with age. Stratified by age, Shock Index was associated with ICU admission (range OR 1.07-1.22) and hospitalisation (range OR 1.06-1.19) (table 3). The discriminative ability for Shock Index was poor for admission to ICU (range AUC 0.59-0.63) or admission to the hospital (range AUC 0.58-0.62) (online supplementary appendix 5). The identified cut-offs per age group had low sensitivity (range 0.27-0.42) and moderate specificity (range 0.79-0.91) for ICU admission. None of

Table 4. Predictive value for different cut-offs of hypotension and/or tachycardia for ICU admission and hospital admission

	ICU admission				Hospital admission			
	Sensitivity (95% CI)	**Specificity (95% CI)**	**Positive likelihood ratio (95% CI)**	**Negative likelihood ratio (95% CI)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Positive likelihood ratio (95% CI)**	**Negative likelihood ratio (95% CI)**
Predictive value for hypotension								
APLS	0.11 (0.09-0.14)	0.96 (0.95-0.96)	2.57 (2.03-3.27)	0.93 (0.90-0.96)	0.06 (0.05-0.07)	0.96 (0.96-0.97)	1.57 (1.32-1.86)	0.98 (0.97-0.99)
PALS / sepsis	0.05 (0.04-0.07)	0.99 (0.99-0.99)	5.06 (3.43-7.46)	0.96 (0.94-0.98)	0.02 (0.02-0.03)	0.99 (0.99-0.99)	2.62 (1.86-3.70)	0.99 (0.98-0.99)
PEWS	0.12 (0.09-0.14)	0.95 (0.95-0.95)	2.38 (1.89-2.99)	0.93 (0.90-0.96)	0.07 (0.06-0.07)	0.95 (0.95-0.96)	1.43 (1.22-1.68)	0.98 (0.97-0.99)
Predictive value for tachycardia								
APLS	0.30 (0.27-0.34)	0.85 (0.84-0.86)	2.02 (1.78-2.29)	0.82 (0.78-0.87)	0.24 (0.23-0.26)	0.89 (0.88-0.89)	2.23 (2.04-2.43)	0.85 (0.83-0.87)
PALS / sepsis	0.25 (0.22-0.28)	0.85 (0.84-0.85)	1.60 (1.39-1.85)	0.89 (0.85-0.93)	0.22 (0.20-0.23)	0.87 (0.87-0.88)	1.71 (1.57-1.87)	0.89 (0.88-0.91)
PEWS	0.55 (0.51-0.59)	0.63 (0.62-0.64)	1.46 (1.36-1.58)	0.72 (0.66-0.79)	0.50 (0.49-0.52)	0.68 (0.67-0.69)	1.58 (1.51-1.66)	0.73 (0.71-0.76)
Predictive value for tachycardia AND hypotension								
APLS	0.05 (0.03-0.07)	0.99 (0.99-0.99)	6.52 (4.26-9.96)	0.96 (0.94-0.98)	0.02 (0.02-0.03)	0.99 (0.99-0.99)	6.54 (4.05-10.6)	0.98 (0.98-0.99)
PALS / sepsis	0.02 (0.01-0.04)	0.99 (0.99-0.99)	11.9 (6.16-23.3)	0.98 (0.97-0.99)	0.01 (0.00-0.01)	0.99 (0.99-0.99)	5.21 (2.35-11.6)	0.99 (0.99-0.99)
PEWS	0.08 (0.06-0.09)	0.98 (0.98-0.98)	4.16 (3.06-5.66)	0.94 (0.92-0.96)	0.04 (0.03-0.04)	0.99 (0.98-0.99)	3.06 (2.35-3.99)	0.97 (0.97-0.98)
Predictive value for tachycardia OR hypotension								
APLS	0.37 (0.33-0.40)	0.81 (0.81-0.82)	1.98 (1.77-2.21)	0.78 (0.73-0.83)	0.28 (0.27-0.29)	0.85 (0.85-0.86)	1.96 (1.81-2.11)	0.84 (0.82-0.86)
PALS / sepsis	0.27 (0.24-0.31)	0.84 (0.83-0.84)	1.69 (1.48-1.93)	0.87 (0.83-0.91)	0.23 (0.22-0.24)	0.87 (0.85-0.87)	1.73 (1.59-1.88)	0.89 (0.87-0.91)
PEWS	0.59 (0.55-0.63)	0.59 (0.59-0.60)	1.45 (1.35-1.56)	0.69 (0.63-0.76)	0.53 (0.51-0.54)	0.65 (0.64-0.66)	1.51 (1.44-1.58)	0.73 (0.69-0.75)

APLS, advanced paediatric life support; ICU, intensive care unit; PALS, Paediatric Advanced Life Support; PEWS, Paediatric Early Warning Score

the identified Shock Index cut-offs had acceptable positive or negative likelihood ratios (online supplementary appendix 6).

The APLS Shock Index cut-off performed similarly with low sensitivity and high specificity (online supplementary appendix 7). The positive likelihood ratio was 3.86 (95%CI 3.1 to 4.8) and negative likelihood ratio was 0.89 (95%CI 0.87 to 0.92).

In febrile children, patients with ill appearance and surgical patients, the hypotension and Shock Index cut-offs showed similar performance. For Shock Index, the highest AUC was found for febrile patients aged >10 years for ICU admission (0.75 95%CI 0.63 to 0.87) (online supplementary appendix 8).

DISCUSSION

In our observational cohort, hypotension has a significant association with serious illness when corrected for tachycardia. However, hypotension showed low sensitivity and high specificity for serious illness in children with routinely measured BP in the ED. The combination of hypotension and tachycardia did not improve the sensitivity further. In addition, although Shock Index was associated with serious illness, acceptable cut-off values could not be identified for different age groups.

Accurate reference values for abnormal vital signs are essential to avoid misclassification. Values based on healthy children may not be accurate for children in the ED, as ill children may present with pain and distress which influences heart rate and BP values. Expert-based cut-offs for low BP are currently used. However, these are not based on large studies and show large variation and are therefore not a good alternative. For example, more than 50% of the children with hypotension according to the APLS were discharged home following the ED visit. Two recent studies presented BP reference ranges and distributions for critically ill children but validated reference values for the paediatric ED population are lacking.[165, 166]

Hypotension is considered a late sign of illness that is preceded by an increase in heart rate. To preserve cardiac output, children compensate by elevating heart rate and systemic vascular resistance. When this compensatory mechanism is inadequate, BP could drop which may indicate shock.[148] Our study showed that heart rate and BP were not correlated. In particular, high Z-scores of heart rate did not correlate with low Z-scores of BP. Moreover, irrespective of tachycardia, cut-offs for hypotension showed a significant association with serious illness.

We focused on tachycardia as this is an early indicator of critical illness and these children could benefit from measuring BP. Bradycardia, however, indicates irreversible

shock. Seriously ill children with bradycardia present with lack of perfusion resulting in cardiopulmonary arrest.[167] Therefore, BP measurement could have limited additional value in children with bradycardia. Furthermore, we did not analyse other predictors of serious illness. In practice, however, heart rate and BP are evaluated with other clinical markers which can be more sensitive predictors for serious illness. Future studies should focus on the combination of BP and other clinical predictors to evaluate the additional value of BP in practice.

Shock Index is associated with mortality in children with septic shock.[153, 154] Research on Shock Index in EDs has mainly focused on injured patients.[155, 156] No reference values exist for the whole age range in children. Acker *et al* proposed age-adjusted cut-offs according to normal vital signs for children >4 year. However, a recent study showed that 2.3% of healthy children had abnormal values according to this definition.[147] Our study found an association between Shock Index and serious illness in different age groups. For children >12 years a 0.1 unit increase in Shock Index relates to odds of 1.22 for ICU admission. However, the discriminative ability for Shock Index was poor. In general, neither of the identified cut-off values had both acceptable sensitivity and specificity.

We focused our analysis on high shock index values to detect severe illness. We acknowledge that low shock index values are also abnormal. Due to the vasopressor response, patients with increased intracranial pressure will have low heart rate and high BP leading to low shock index values.

Although hypotension showed high specificity for serious illness, the sensitivity was very low, regardless of the used definition. The combination of hypotension and tachycardia did not improve the sensitivity or the specificity for predicting serious illness. PALS[158] had good rule-in value having good specificity and high positive likelihood ratios. However, for early recognition of severely ill children in the ED, it is important to rule out serious illness. Hypotension and tachycardia lack these characteristics, having low sensitivity and poor negative likelihood ratios for serious illness. Considering that accurate BP measurement is time consuming for nurses[168], these results suggest limited value of routine BP measurement in all children attending the ED.

Strengths of this study are the use of three hypotension cut-offs that are widely used in clinical practice. In addition, our analyses were based on a large cohort of paediatric ED patients of all ages with different presenting problems. We used routine data and therefore our results are representative of clinical practice.

This study has some limitations. First, patients were included when BP and heart rate were measured. This selected group is more severely ill, comprising older children, more

highly urgent cases and more ICU admissions. This could potentially bias our findings. However, this reflects measurement of BP in the practice of the ED. The frequency of BP investigation and the increase with age and urgency was similar to previous studies.[26, 145] In addition, the population of our tertiary university hospital consists of more children with comorbidities and more severely ill children. In settings with low prevalence of serious illness, less yield could be expected. Second, we used hospital admission and ICU admission to define serious illness. These outcomes are widely used in literature and applicable to large datasets.[160, 169, 170] As reasons for ICU admission following ED visit include life-threatening conditions, the presence of hypotension could have influenced the decision for ICU admission. Hospital admission could occur for various conditions as fractures or bronchiolitis which are unlikely to develop low BP. Furthermore, accurate measurement of BP in children in the ED is challenging. Movement of limbs and uncooperativeness interfere with the measurements. Moreover, the correct cuff size and technique need to be applied. Therefore, the quality of BP measurement should be taken into account.

Finally, our study aimed to evaluate the value of routine BP measurements in children for the recognition of serious illness. We acknowledge that BP measurement may be indicated in the ED for diagnostics, detection of hypertension, follow-up or therapy monitoring.

CONCLUSION

Our observational study demonstrates that hypotension is associated with serious illness, independent of heart rate. Although the specificity of hypotension is high, the sensitivity for serious illness is very low. The combination of hypotension and tachycardia did not further improve the sensitivity. Shock index is related to serious illness, however we could not identify acceptable cut-off values. These findings suggest limited value of measuring routine BP to detect hypotension in all attending children. Future studies need to investigate which specific patients could benefit from BP measurement and should focus on developing accurate reference values for hypotension and Shock Index that are applicable in the ED.

SUPPORTING INFORMATION

Available as online web appendix on the website of the Archive of Disease in Childhood:
- Appendix 1. Definitions of tachycardia
- Appendix 2. Definition of high shock index APLS
- Appendix 3. Characteristics patients with no blood pressure/heart rate measurement
- Appendix 8 Results of subgroup analysis in children with fever, ill appearance, surgical

Appendix 4. Frequencies of hypotension and tachycardia

Table 4.1 Frequencies of hypotension and tachycardia according to 3 different cut-offs (N=10,698)

Hypotension; n (%)	
APLS	504 (4.7)
PALS/septic shock screening tool	133 (1.2)
PEWS by Parshuram	571 (5.3)
Tachycardia; n (%)	
APLS	1692 (15.8)
PALS/septic shock screening tool	1709 (18.0)
PEWS by Parshuram	4113 (38.4)

APLS, Advanced Paediatric Life Support; PALS, Paediatric Advanced Life Support; PEWS, Paediatric Early Warning Score

Appendix 5 – Area under the curve (AUC) for Shock Index

Table 5.1 Area under the curve (AUC) for the receiver operating characteristics with 95%CI for Shock Index

	ICU admission	**Hospital admission**
All ages	0.63 (0.60-0.65)	0.63 (0.60-0.65)
0-2 years (n=1588)	0.63 (0.58-0.67)	0.62 (0.59-0.64)
2-10 years (n=5011)	0.56 (0.52-0.60)	0.58 (0.56-0.59)
10-16 years (n=4099)	0.59 (0.54-0.64)	0.58 (0.56-0.59)

ICU, Intensive care unit

Appendix 6. Predictive value for high cut-offs of Shock Index

Table 6.1 Predictive value for cut-offs of Shock Index according to different age groups

ICU admission	Shock Index Cut-off*	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
0-2 years (n=1588)	1.63	0.42 (0.35-0.49)	0.79 (0.77-0.81)	2.00 (1.65-2.44)	0.74 (0.65-0.83)
2-10 years (n=5011)	1.18	0.36 (0.79-0.81)	0.80 (0.79-0.81)	1.83 (1.53-2.18)	0.79 (0.72-0.88)
10-16 years (n=4099)	1.05	0.27 (0.21-0.34)	0.91 (0.91-0.92)	3.16 (2.45-4.07)	0.79 (0.73-0.87)
Hospital admission	**Shock Index Cut-off***	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Positive likelihood ratio (95% CI)**	**Negative likelihood ratio (95% CI)**
0-2 years (n=1588)	1.45	0.48 (0.45-0.51)	0.72 (0.69-0.76)	1.74 (1.52-1.99)	0.72 (0.67-0.78)
2-10 years (n=5011)	1.13	0.34 (0.32-0.37)	0.79 (0.77-0.80)	1.63 (1.49-1.79)	0.83 (0.80-0.86)
10-16 years (n=4099)	0.91	0.30 (0.28-0.33)	0.82 (0.81-0.84)	1.73 (1.54-1.94)	0.84 (0.81-0.88)

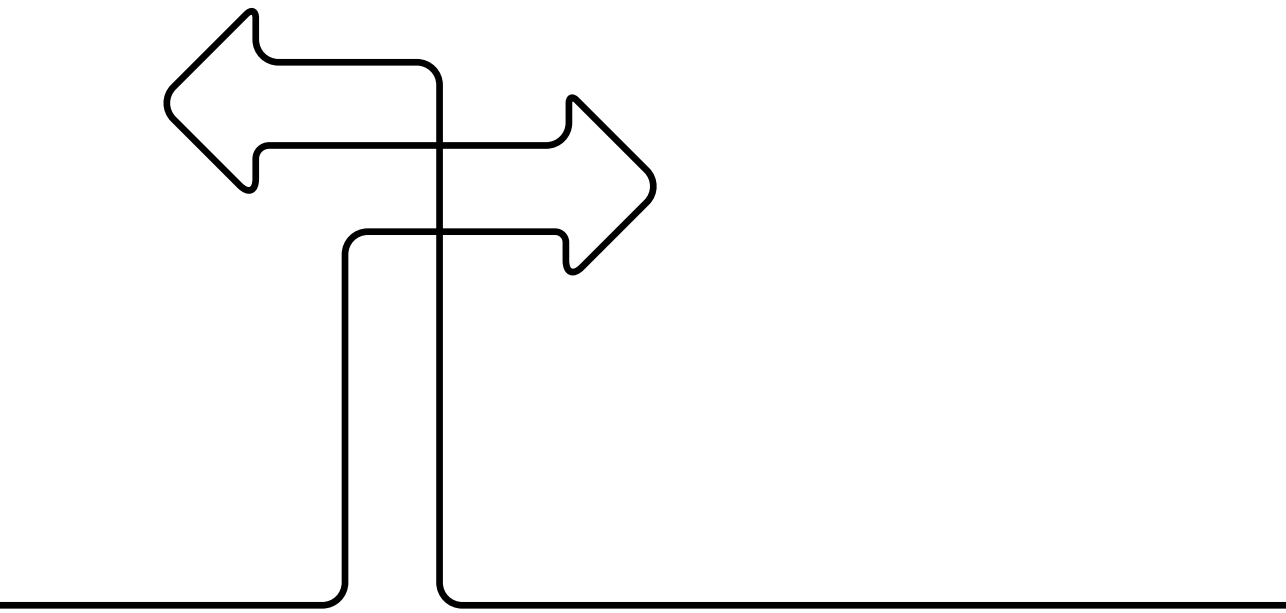
*Cut-off value determined by Youden's index
ICU, intensive care unit

Appendix 7. Predictive value of high Shock Index: cut-off defined by APLS abnormal vital signs

Table 7.1 Predictive value of high shock index cut-off by APLS*

ICU admission	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Positive likelihood ratio (95% CI)
Total	0.14 (0.11-0.17)	0.96 (0.96-0.97)	3.86 (3.1-4.8)	0.89 (0.87-0.92)
Ill appearance	0.22 (0.17-0.28)	0.92 (0.90-0.93)	2.71 (1.95-3.75)	0.85 (0.79-0.92)
Fever	0.20 (0.14-0.28)	0.92 (0.91-0.93)	2.49 (1.71-3.65)	0.87 (0.79-0.95)
Surgical	0.04 (0.02-0.08)	0.99 (0.99-0.00)	4.80 (1.74-13.3)	0.97 (0.94-1.0)
Hospital admission	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Positive likelihood ratio (95% CI)**	**Positive likelihood ratio (95% CI)**
Total	0.08 (0.07-0.08)	0.98 (0.97-0.98)	3.33 (2.75-4.03)	0.95 (0.94-0.96)
Ill appearance	0.13 (0.11-0.15)	0.96 (0.94-0.97)	3.03 (1.95-4.72)	0.91 (0.88-0.94)
Fever	0.15 (0.13-0.17)	0.95 (0.94-0.96)	3.23 (2.48-4.22)	0.89 (0.87-0.92)
Surgical	0.02 (0.01-0.03)	0.99 (0.99-0.99)	2.79 (1.05-7.41)	0.99 (0.98-1)

*APLS cut-off calculated by dividing APLS tachycardia value to APLS hypotension value for each age group
ICU, Intensive care unit



Chapter 9

A comparison of clinical paediatric guidelines for hypotension with population-based lower centiles: a systematic review

Nienke N Hagedoorn, Joany M Zachariasse, Henriëtte A Moll

Crit Care. 2019 Nov 27;23(1):380

ABSTRACT

Background Different definitions exist for hypotension in children. In this study, we aim to identify evidence-based reference values for low blood pressure and to compare these with existing definitions for systolic hypotension.

Methods We searched online databases until February 2019 (including MEDLINE, EMBASE, Web of Science) using a comprehensive search strategy to identify studies that defined age-related centiles (first to fifth centile) for non-invasive systolic blood pressure in healthy children < 18 years. Existing cut-offs for hypotension were identified in international guidelines and textbooks. The age-related centiles and clinical cut-offs were compared and visualized using step charts.

Results Fourteen studies with population-based centiles were selected, of which 2 addressed children < 1 year. Values for the fifth centile differed 8 to 17 mmHg for age. We identified 13 clinical cut-offs of which only 5 reported accurate references. Age-related cut-offs for hypotension showed large variability (ranging from 15 to 30 mmHg). The clinical cut-offs varied in agreement with the low centiles. The definition from Paediatric Advanced Life Support agreed well for children < 12 years but was below the fifth centiles for children > 12 years. For children > 12 years, the definition of Parshuram's early warning score agreed well, but the Advanced Paediatric Life Support definition was above the fifth centiles.

Conclusions The different clinical guidelines for low blood pressure show large variability and low to moderate agreement with population-based lower centiles. For children < 12 years, the Paediatric Advanced Life Support definition fits best but it underestimates hypotension in older children. For children > 12 years, the Advanced Paediatric Life Support overestimates hypotension but Parshuram's cut-off for hypotension in the early warning score agrees well. Future studies should focus on developing reference values for hypotension for acutely ill children.

INTRODUCTION

Vital signs are important in the recognition of acutely ill children. One parameter associated with serious illness is hypotension.[149, 171, 172] Because normal blood pressure values vary with age, accurate age-related reference values are needed to correctly identify hypotension in children and guide interventions.

Blood pressure can be measured by invasive, oscillometric and auscultatory methods. In addition, various outcome measures for blood pressure exist such as mean arterial pressure, and diastolic and systolic blood pressure. Paediatric guidelines propose different definitions of hypotension and in general use cut-off values of systolic blood pressure.[28, 146, 173] Although not based on evidence, several guidelines use the fifth percentile of systolic blood pressure in healthy children as cut-off for hypotension.[28, 158, 174] Moreover, it is unclear how well these guidelines discriminate between normal and low blood pressure. To date, no study has summarized the available evidence on reference values of low systolic blood pressure in children.

This study aims to identify population-based reference values for non-invasive low blood pressure in healthy children and to compare these with cut-offs for hypotension defined by existing paediatric guidelines.

METHODS

Search strategy and selection of population-based studies

We systematically searched databases including MEDLINE, EMBASE and other databases (1950 to 14 February 2019) to identify primary studies that defined lower centiles for non-invasive systolic blood pressure measurement in healthy children (Additional file 1: detailed search strategy). Studies that were included were published in English, recorded blood pressure and defined age-related centiles for systolic blood pressure (first to fifth centile) on a minimum of 100 children aged < 18 years. Studies were excluded if populations involved children with underlying diseases, or studies reporting on premature neonates, measurements during anaesthesia, exercise or orthostasis. We excluded populations from low- and middle-income countries since factors influencing blood pressure levels such as body composition and nutrition, are different compared to high-income countries.[175] We excluded abstracts, reviews and commentaries, and studies reporting on lower centiles solely derived from mathematical analysis. One researcher (NH) conducted the first selection, and two researchers (NH, JZ) independently conducted the second and third selection. Disagreements were discussed and agreed upon consensus or discussed with a third researcher (HM) for majority decision.

Data extraction and analysis

For the selected studies, data were extracted by one researcher (NH) and included country, population, setting, sample size, age range, blood pressure measurement method and age-specific centiles (P1-P5). We included the centiles for non-overweight children and for the median height if blood pressure centile values were reported for different height categories. The age-specific fifth centiles were summarized using weighted medians and interquartile ranges for age categories which involved three or more studies. If sample sizes were only given for age ranges > 1 year, we estimated the sample size per age group by dividing the total sample size by the number of years.

Quality assessment

No specific tool exists for quality assessment of observational studies.[176] The Quality Assessment of Diagnostic Accuracy Studies-2 checklist was the most appropriate to use for these observational studies.[78] This checklist covers risk of bias and applicability judgments on four domains: patient selection, index test, reference standard and flow and timing. For each question, studies were classified as high, low or unclear. Disagreements were agreed upon consensus.

Cut-off values for hypotension from clinical guidelines

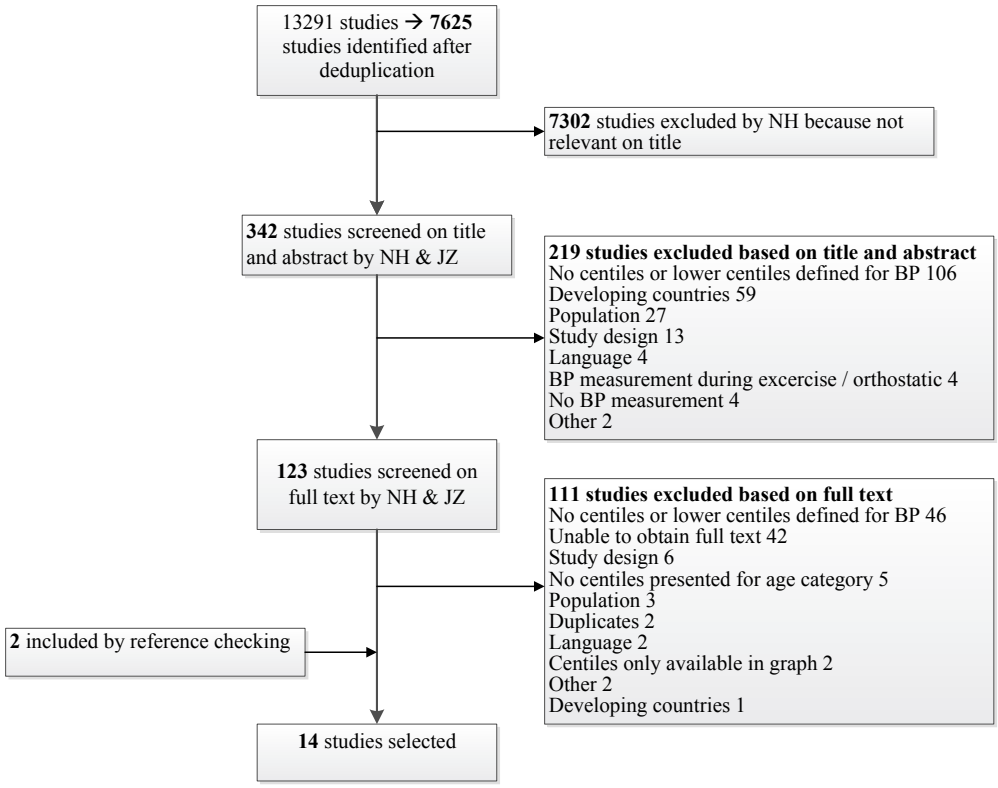
We selected a sample of clinical cut-offs for hypotension by consulting experts, well-known textbooks and resuscitation, emergency care and sepsis guidelines. Clinical cut-offs included recommended target values for hypotension defined by systolic blood pressure. For each clinical cut-off, we determined the presence of a literature reference and whether this reference agreed with the cut-off values. To compare clinical cut-offs with the population-based centiles identified in the literature, we plotted the age-specific fifth centile values in a step chart separate for boys and girls. Data analyses were performed in SPSS version 25.0 and R version 3.4.

RESULTS

Population-based studies

Our systematic search identified 7625 studies. After the study selection process, we included 14 studies in the final selection that defined lower centiles for non-invasive systolic blood pressure measurement in healthy children (Fig. 1).

Fig. 1 Study selection process



BP, blood pressure

The median samples size was 5362 (IQR 1760-11,940). Seven out of 14 studies used an automatic oscillometric device for blood pressure measurement. Two studies included children aged < 1 year (Table 1). Studies included populations from Europe (*n* = 8), North America (*n* = 3), Australia (*n* =2) and Asia (*n* = 1). Four studies excluded overweight patients. For development of the centiles, 11 studies used the average of multiple blood pressure measurements and 3 studies used only the first measurement. Blood pressure centiles were stratified by gender (*n* = 12), height (*n* = 4), ethnicity (*n* = 1) and overweight vs non-overweight (*n* = 2). Studies most frequently reported the fifth centile (*n* = 13), in which the third centile (*n* = 2) and first centile (*n* = 3) were also reported separately. One study only reported the first and third centiles.

The fifth centiles of the population-based studies showed variation ranging across the age groups from 7 to 17 mmHg for boys (Fig. 2) and 7 to 22 mmHg for girls (Additional file 2). Median values and interquartile ranges of the lower fifth centiles are provided in Additional files 3 and 4.

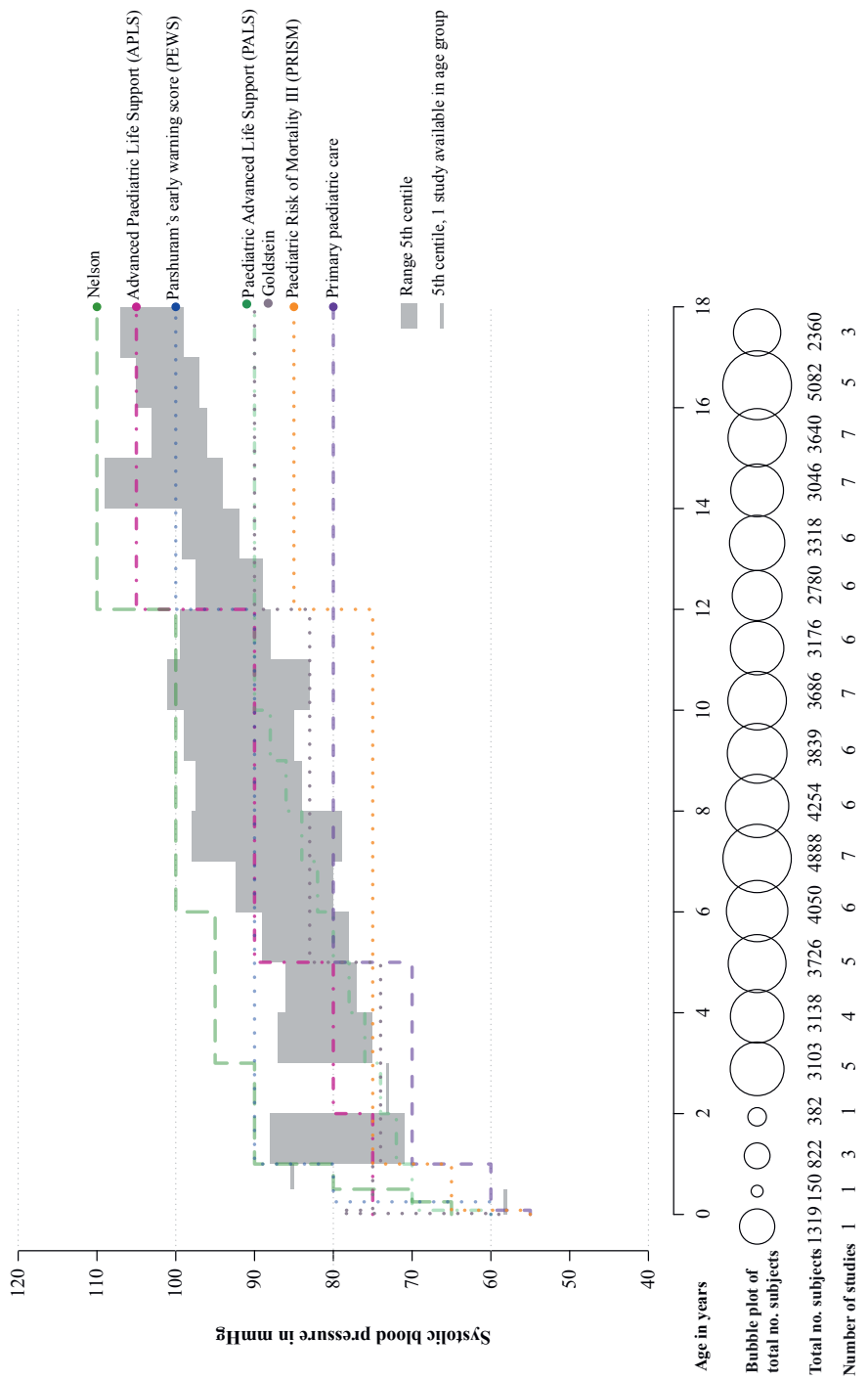


Fig. 2 Clinical definitions for hypotension and range of 5th centile of systolic blood pressure for boys

Table 1. Characteristics of included studies

Author	Country	Inclusion	Exclusion	Age range (yrs)	Setting	Sample size	Method of measurement	Defined BP centiles	Determinants of age-specified centiles	Measurement used for analysis	Main outcome
Antal *et al.* (2004) [177]	Hungary	Secondary school	Using anti-hypertensive medication	15-18	Community setting	6345	Oscill.	P3, P5	Sex	first measurement	Assessment of age- and gender-specific anthropometric parameters and blood pressure values
Barba *et al.* (2014) [178]	8 EU countries	Non-overweight children	Overweight	2-10.9	Unspecified	13,547	Oscill.	P1, P3	Sex, height	Mean of first and second measurement	Provide oscillometric blood pressure reference values
Blake *et al.* (2000) [179]	Australia	Cohort from a tertiary perinatal centre. Follow-up at age 1, 3 and 6 years	x	1-6	Unspecified	2876	Oscill.	P5	Sex	Mean of two measurements	To develop age- and gender-specific reference ranges for BP
Grajda et al. (2017) [180]	Poland	Healthy pre-school children	Congenital, chronic or acute disorders and medication affecting growth or BP levels.	3-6	Community setting	4378	Oscill.	P1, P5	Sex, height	Mean of second and third measurement	To develop age- and gender-specific ranges for BP in pre-school children
Hediger *et al.* (1984) [181]	USA	Black adolescents	x	11-17	Unspecified	621	Auscul.	P5	Sex	Mean of two measurements	Percentiles for black adolescents for resting BP and 60-second pulse rate
Kent *et al.* (2007) [182]	Australia	Term infants	Congenital anomalies, birth weight <3rd percentile, sepsis, NICU admission. Maternal hypertension, diabetes, use of illicit substances.	0 -1	Hospital: postnatal clinical, other in a non-clinical room	406	Oscill.	P5	X	Mean of three measurements	Normative BP during first year of life of healthy infants

Karmar *et al.* (2014) [183]	Sweden	Children, junior school	Physical health problems, medication that affects BP	6-16	Community setting	1470	Oscill.	P5	Sex	Mean of second and third measurement	Cross-sectional normative casual BP standards
Krzyzaniak *et al.* (2009) [184]	Poland	School children	x	7-18	Community setting	6447	Auscul.	P5	Sex, height	Mean of two measurements on three different days	To develop age- and gender-specific reference ranges
Lurbe *et al.* (1994) [185]	Spain	Normotensive children	Systemic and renal disease	6-16	Primary care	248	Oscill.	P5	Sex, casual and ambulatory BP	Mean of three measurements and means of daytime measurements	Assess reference values of ambulatory blood pressure
Rosner *et al.*(2008) [186]	USA	11 large pediatric blood pressure studies (based on Pediatric Task Force database)[187]	Overweight	1-17	Unspecified	36,914	Auscul.	P1,P5	Sex, height	first measurement	Norms for childhood BP among normal-weight children
Sarganas (2018) [147]	Germany	Healthy children and adolescents	Chronic conditions or medication influencing growth or BP. Overweight (BMI>90th centile).	3-17	Community setting	14,836	Oscill.	P1, P5	Sex, height	Mean of two measurements	Fifth percentile of BP according to age, sex and height.
Satoh et al. (2016) [188]	Japan	Full-term singleton newborns	Twin newborns, miscellaneous abnormalities, missing Apgar score, condition during BP measurement	0	Hospital	2628	Oscill.	P5	Sex	first measurement	Estimate BP and pulse rate in healthy newborns
Schwandt et al. (2015) [189]	Germany	German parents	Metabolic, cardiovascular, endocrine, malignant disorder, specific medication, non-German ethnicity.	3-18	Community setting	22,051	Auscul.	P3, P5	Sex, overweight and non-overweight	Mean of two measurements	Develop auscultatory BP growth charts
Weiss *et al.* (1973) [190]	USA	Non-institutionalized children	x	6-11	Hospital: one visit	7119	Auscul.	P5	Sex, race	Mean of two measurements	Distribution of BP level 6-11 years

Auscul., auscultatory; BP, blood pressure; EU, European Union; NICU, Neonatal intensive care unit; Oscill, oscillometric; P1, First centile; P3, Third centile; P5, fifth centile; USA, United states of America

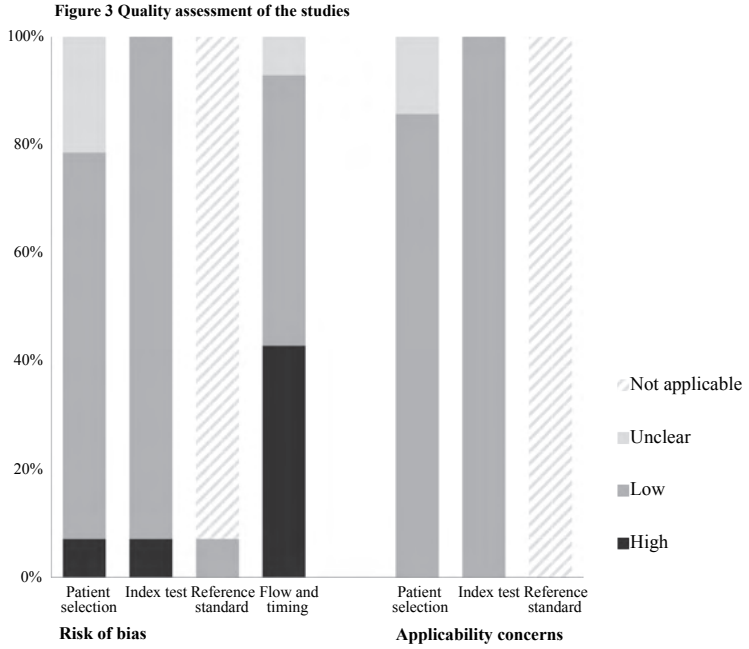
Quality of the population studies was generally good. No concerns regarding applicability were found in 12 out of 14 studies. Six studies had high risk of bias in the patient flow and timing domain, due to poor reporting of how missing data were handled (Table 2, Fig. 3).

Table 2. Quality Assessment of the studies

	Risk of bias				Applicability concerns		
	Patient selection	**Index test**	**Reference standard**	**Flow and timing**	**Patient selection**	**Index test**	**Reference standard**
Antal[177]	Low	Low	n/a	Unclear	Low	Low	n/a
Barba[178]	Low	Low	n/a	High	Low	Low	n/a
Blake[179]	Low	Low	n/a	High	Low	Low	n/a
Grajda[180]	Low	Low	n/a	Low	Low	Low	n/a
Hediger[181]	Low	Low	n/a	High	Low	Low	n/a
Kent[182]	Low	Low	n/a	High	Low	Low	n/a
Karmar[183]	Low	Low	n/a	High	Low	Low	n/a
Krzyzaniak[184]	Unclear	Low	n/a	Low	Low	Low	n/a
Lurbe[185]	High	Low	Low	Low	Low	Low	n/a
Rosner[186]	Unclear	High	n/a	Low	Unclear	Low	n/a
Sarganas[147]	Low	Low	n/a	Low	Low	Low	n/a
Satoh[188]	Unclear	Low	n/a	High	Unclear	Low	n/a
Schwandt[189]	Low	Low	n/a	Low	Low	Low	n/a
Weiss[190]	Low	Low	n/a	Low	Low	Low	n/a

n/a: not applicable

Fig. 3 Quality assessment of the studies



Cut-off values for hypotension from clinical guidelines

We identified 13 clinical cut-offs for hypotension of which 8 referred to a literature reference (Additional file 5). Five cut-offs provided an accurate literature reference [158, 159, 191-193], of which four out of five referred to the fifth centile of healthy children. In two textbooks, the values of the literature reference did not agree with the provided cut-offs.[194, 195] One literature reference could not be obtained.[196] Age-specific cut-off values for hypotension showed large differences, ranging from 15 to30 mmHg (Fig. 2, Additional file 5).

Comparison of population-based studies with cut-off values for hypotension from clinical guidelines

The clinical hypotension cut-offs showed poor to moderate agreement with the lower centiles derived from population-based studies (Fig. 2). The frequently used hypotension cut-off from Advanced Paediatric Life Support (APLS)[173] showed moderate agreement for children < 12 years, but was above the highest fifth centile values for children > 12 years. The cut-off from Paediatric Advanced Life Support (PALS) agreed well for children < 12 years but was below the fifth centile values for children > 12 years. The cut-off of Parshuram's early warning score (PEWS) agreed well for children > 12 years.[140] Three other cut-offs were mostly below the fifth centiles (Goldstein, primary paediatric care and Paediatric Risk of Mortality III (PRISM III))[40, 193, 194] and one cut-off had higher values (Nelson).[197]

DISCUSSION

This systematic review demonstrates large variation among commonly used paediatric reference values for systolic hypotension. In general, the clinical guidelines are not based on available evidence and showed variable agreement with existing population-based blood pressure centiles. The reviewed literature addressing population-based centiles showed limited studies in children < 1 year of age.

Reference ranges of blood pressure are influenced by multiple factors such as age, gender, height, ethnicity and method of measurement.[187] In the literature, low centiles for blood pressure are often presented for different ages and in some cases for height. To facilitate interpretation, guidelines provide simplified cut-off values for hypotension for various age groups. For early recognition of acutely ill children, these simplified reference values are essential for clinicians.

The evidence for clinically used cut-offs for hypotension is mostly unclear as only five clinical cut-offs for hypotension reported accurate literature references. Our systematic

search shows availability of population-based centiles that could provide evidence for lower reference values of blood pressure. Although not evidence based, we propose that clinical cut-offs for hypotension should not exceed the fifth centile. Clinical cut-offs that are generally below the fifth centile may possibly be too low, whilst clinical cut-offs that are generally above the fifth centile may be too high. These high clinical cut-offs may classify too many patients incorrectly as hypotensive since by definition 5% of healthy children will fall below this centile. In children < 12 years the values of PALS have good agreement with the low centiles, but for children age > 12 years the PALS could possibly be too low.

Our results are in line with a previous study that compared three clinical cut-offs with the fifth centile, based on a mathematical analysis of a large sample of healthy children.[28] They reported that the fifth centile for systolic blood pressure was generally below three clinical cut-offs for hypotension. Sarganas et al. found that low centiles from a German and US population were higher than the PALS definition in children > 13 years.[147] In contrast to the previous studies, our study conducted an exhaustive systematic search for population-based centiles in all ages and compared them with a large sample of cut-offs for hypotension that are widely used in clinical practice. Our study identified only two studies that provided blood pressure centiles in children < 1 year including one study in new-borns and one at age of 6 months. [182, 188] Therefore, more studies providing reference values of blood pressure in children < 1 year are required.

Reference values based on healthy children may not be accurate for acutely ill children, as pain and distress could increase blood pressure values. In addition, cuff size, movement of limbs, crying and uncooperativeness influence the measured values. In the interpretation of the measured values, these factors should be accounted for.

There is no consensus on which definition of hypotension should be used for the assessment of acutely ill children. Hypotension defined by APLS, PALS and PEWS, showed an association with serious illness, adjusted for tachycardia. These definitions, however, lacked sensitivity for serious illness.[172] In our systematic review, the PALS cut-off showed the best agreement with the values based on healthy children with an average of 4 mmHg difference from the weighted median of the population-based fifth centiles. In addition, current guidelines do not agree on treatment targets for blood pressure after identification of hypotension in critically ill children. The goal for treatment target of blood pressure is to maintain adequate tissue perfusion. The guideline of International Liaison Committee on Resuscitation recommends targeting systolic blood pressure values higher than the fifth percentile for children who are post-cardiac arrest, [198] whilst the APLS and the surviving sepsis campaign[149] advise to maintain normal blood pressure for age without defining specific measures. The American College of Critical Care Medicine

recommends to use the 50th centile of the mean arterial pressure (MAP) and to use perfusion pressure (MAP- central venous pressure) to guide treatment.[159] Some evidence is available suggesting higher MAP levels are needed to improve outcome in traumatic brain injury and central nervous system infections in children.[171, 199] Trials in adult critically ill patients with septic shock showed that targeting higher mean arterial pressure levels of 75-85 mmHg did not influence mortality or other adverse events.[200, 201] Future trials will need to evaluate different blood pressure measures and targets in acutely ill children and relate those to interventions and relevant clinical outcomes.

Our review focused on systolic blood pressure and did not include mean arterial blood pressure or diastolic blood pressure. Although the mean arterial pressure is often used in critical care, we focused on systolic hypotension for general illness, since in general, clinical guidelines only report hypotension definitions of systolic blood pressure.

Strengths and limitations

Major strengths of this study are the use of an extensive search strategy, the overview of low reference values of blood pressure in healthy children covering all ages and the comparison with a diverse sample of clinical cut-offs of hypotension that are widely used in practice. Although we used a sensitive search strategy in multiple databases, it is possible we have not included all available data. Since we focused on lower age-related centiles, we excluded studies that reported blood pressure centiles solely for height or body mass index.

This study has some limitations. First, the selected sample of clinical definitions was not exhaustive and various blood pressure cut-offs in early warning scores and mortality scores were not included. We selected Parshuram's early warning score and the PRISM III mortality score as these have been validated and are commonly used in practice. We acknowledge that these cut-offs are part of a score containing other clinical markers. In addition, the PRISM III score has been developed specifically for predicting mortality in critically ill children.

Second, blood pressure is determined by height and we only included blood pressure values for the median height value. However, height is usually not available in the assessment of acutely ill children and none of the clinical guidelines accounted for height. Third, we focused on non-invasive measurement methods including oscillometric and auscultatory measurements. Oscillometric measured values could be different than auscultatory measurements.[202] As different devices were used in the studies and their validity in assessment of low blood pressure is unknown, we combined centiles for oscillometric and auscultatory measurements. Fourth, since non-invasive blood pressure measurements could overestimate hypotension when compared to invasive arterial measurement, generalization of our study to invasive measurements should be undertaken with caution. [203-205]

CONCLUSION

Large variation exists among paediatric cut-offs for hypotension. In general, these clinical definitions are not evidence-based and have variable agreement with existing population-based blood pressure lower centiles.

For children < 12 years, the PALS definition agreed well. For children > 12 years, the PEWS agreed well but the PALS cut-off possibly underestimates and the APLS overestimates hypotension. Future studies should focus on developing reference values for hypotension for acutely ill children.

ACKNOWLEDGEMENTS

We would like to thank Wichor M. Bramer, Medical Library Erasmus MC, for the development of the search strategy.

SUPPLEMENTARY INFORMATION

Available as online web appendix on the website of Critical Care:
- Additional file 5. Clinical cut-offs for hypotension

Additional file 1. Systematic search strategy

Table 1.1 Systematic search strategy

Database	Number of references	After deduplication
Embase.com	5225	5089
Medline (ovid)	4854	1688
Web-of-science	2725	585
Cochrane	141	23
Cinahl (ebsco)	21	12
Lilacs	71	57
Scielo	25	3
Proquest	29	24
Google scholar	200	144
Total	**13291**	**7625**

Embase.com

((('blood pressure'/de OR 'blood pressure measurement'/exp OR 'blood pressure monitoring'/exp OR 'blood pressure variability'/exp) AND ('statistical analysis'/de OR 'statistical distribution'/exp OR statistics/exp)) OR (normotension* OR ((norm* OR healthy OR population OR nomogram* OR curve* OR centile* OR survey* OR distribut* OR statistic* OR trend* OR differen* OR varia* OR 'z score' OR reference* OR standard*) NEAR/9 ('blood pressure' OR 'blood pressures' OR bp))):ab,ti) AND (child/exp OR newborn/exp OR adolescent/exp OR adolescence/exp OR (adolescen* OR infan* OR newborn* OR (new NEXT/1 born*) OR child* OR pediatric* OR paediatric*):ab,ti) AND ('cohort analysis'/exp OR 'population research'/exp OR 'population group'/de OR 'cross-sectional study'/exp OR 'longitudinal study'/exp OR population/de OR (cohort* OR population* OR (cross NEXT/1 section*) OR longitudinal*):ab,ti)

Medline (ovid)

(((exp "blood pressure"/ OR exp "Blood Pressure Determination"/) AND ("Statistics as Topic"/ OR exp "Statistical Distributions"/ OR statistics/)) OR (normotension* OR ((norm* OR healthy OR population OR nomogram* OR curve* OR centile* OR survey* OR distribut* OR statistic* OR trend* OR differen* OR varia* OR "z score" OR reference* OR standard*) ADJ9 ("blood pressure" OR "blood pressures" OR bp))).ab,ti.) AND (exp child/ OR exp infant/ OR adolescent/ OR exp pediatrics/ OR (adolescen* OR infan* OR newborn* OR

(new ADJ born*) OR child* OR pediatric* OR paediatric*).ab,ti.) AND ("Cohort Studies"/ OR "Population Groups"/ OR "Cross-Sectional Studies"/ OR "Longitudinal Studies"/ OR population/ OR (cohort* OR population* OR (cross ADJ section*) OR longitudinal*).ab,ti.)

Web-of-science

TS=(((normotension* OR ((norm* OR healthy OR population OR nomogram* OR curve* OR centile* OR survey* OR distribut* OR statistic* OR trend* OR differen* OR varia* OR "z score" OR reference* OR standard*) NEAR/9 ("blood pressure" OR "blood pressures" OR bp)))) AND ((adolescen* OR infan* OR newborn* OR (new NEAR/1 born*) OR child* OR pediatric* OR paediatric*)) AND ((cohort* OR population* OR (cross NEAR/1 section*) OR longitudinal*)))

Cochrane

((normotension* OR ((norm* OR healthy OR population OR nomogram* OR curve* OR centile* OR survey* OR distribut* OR statistic* OR trend* OR differen* OR varia* OR 'z score' OR reference* OR standard*) NEAR/9 ('blood pressure' OR 'blood pressures' OR bp))):ab,ti) AND ((adolescen* OR infan* OR newborn* OR (new NEXT/1 born*) OR child* OR pediatric* OR paediatric*):ab,ti) AND ((cohort* OR population* OR (cross NEXT/1 section*) OR longitudinal*):ab,ti)

Cinahl (ebsco)

(((MH "blood pressure+" OR MH "Blood Pressure Determination+" OR MH "Blood Pressure Devices+") AND (MH "Statistics")) OR SU (normotension* OR ((norm* OR healthy OR population OR nomogram* OR curve* OR centile* OR survey* OR distribut* OR statistic* OR trend* OR differen* OR varia* OR "z score" OR reference* OR standard*) N3 ("blood pressure" OR "blood pressures" OR bp)))) AND (MH child+ OR MH infant+ OR adolescent+ OR MH pediatrics+ OR SU (adolescen* OR infan* OR newborn* OR (new N1 born*) OR child* OR pediatric* OR paediatric*)) AND (MH "Cross-Sectional Studies+" OR MH population+ OR SU (cohort* OR population* OR (cross N1 section*) OR longitudinal*))

Pubmed publisher

(((("blood pressure"[mh] OR "Blood Pressure Determination"[mh]) AND ("Statistics as Topic"[mh] OR "Statistical Distributions"[mh])) OR (normotension*[tiab] OR ((norm[tiab] OR norms[tiab] OR normal*[tiab] OR healthy OR population OR nomogram*[tiab] OR curve*[tiab] OR centile*[tiab] OR survey*[tiab] OR distribut*[tiab] OR statistic*[tiab] OR trend*[tiab] OR differen*[tiab] OR varia*[tiab] OR "z score" OR reference*[tiab] OR standard*[tiab]) AND ("blood pressure" OR "blood pressures" OR bp)))) AND (child[mh] OR infant[mh] OR adolescent[mh] OR pediatrics[mh] OR (adolescen*[tiab] OR infan*[tiab] OR newborn*[tiab] OR (new born*[tiab]) OR child*[tiab] OR pediatric*[tiab] OR paediatric*[tiab])) AND ("Cohort Studies"[mh] OR "Population Groups"[mh] OR

"Cross-Sectional Studies"[mh] OR "Longitudinal Studies"[mh] OR population[mh] OR (cohort*[tiab] OR population*[tiab] OR (cross section*[tiab]) OR longitudinal*[tiab])) AND publisher[sb]

Google scholar

normotension|normotensive|"normal|healthy|population|standard blood pressure" adolescents|adolescence|infants|infancy|newborn|children|pediatric|paediatric cohort|cohorts|population|"cross section|sectional"|longitudinal

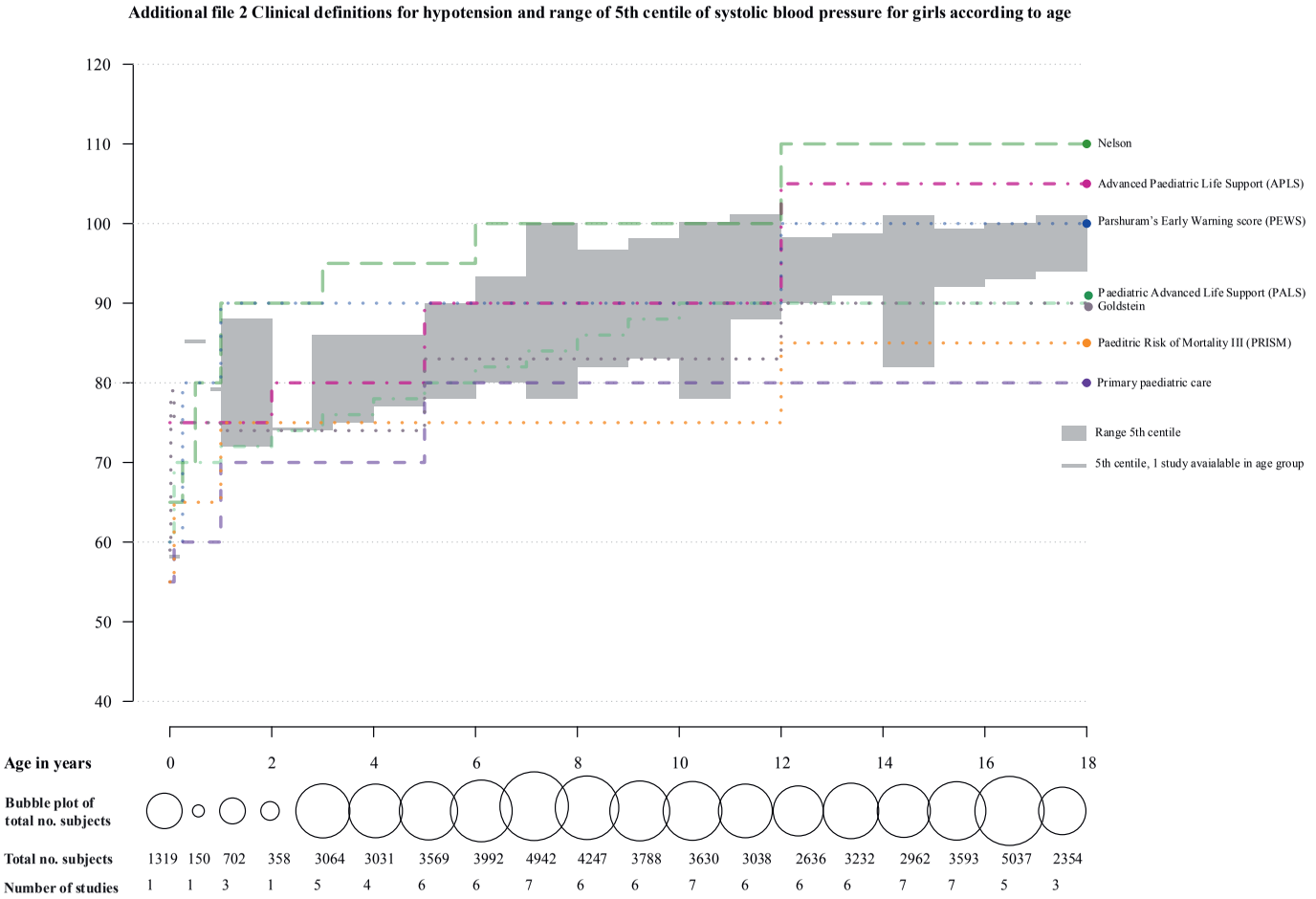
lilacs
scielo

(normotens* OR "normal blood pressure" OR "healthy blood pressure" OR "population blood pressure" OR "standard blood pressure") AND (adolescen* OR infan* OR newborn OR child* OR pediatric* OR paediatric*) AND (cohort* OR population OR "cross section" OR "cross sectional" OR longitudinal)

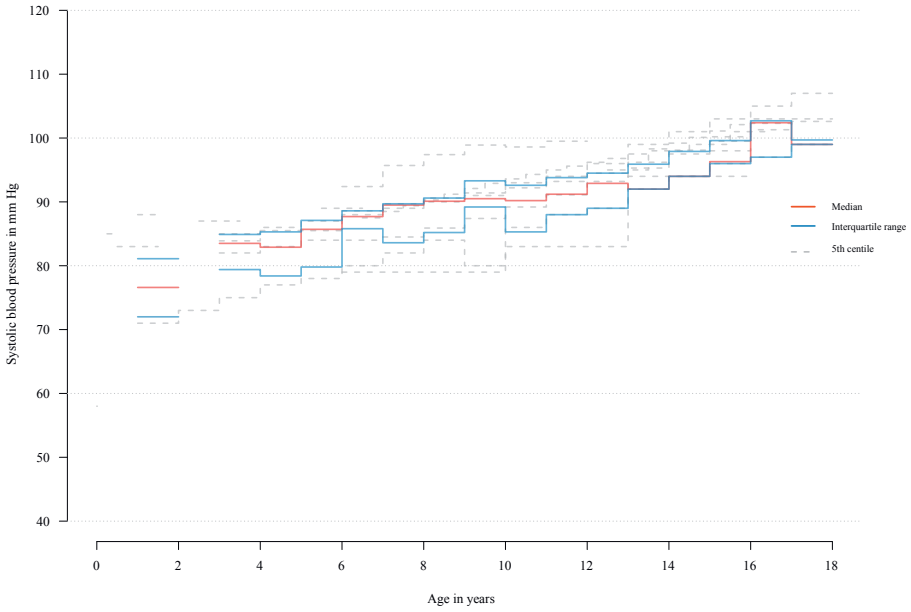
Proquest

(ti(normotens* OR "normal blood pressure" OR "healthy blood pressure" OR "population blood pressure" OR "standard blood pressure") OR ab(normotens* OR "normal blood pressure" OR "healthy blood pressure" OR "population blood pressure" OR "standard blood pressure")) AND (ti(adolescen* OR infan* OR newborn OR child* OR pediatric* OR paediatric*) OR ab(adolescen* OR infan* OR newborn OR child* OR pediatric* OR paediatric*)) AND (ti(cohort* OR population OR "cross section" OR "cross sectional" OR longitudinal) OR ab(cohort* OR population OR "cross section" OR "cross sectional" OR longitudinal))

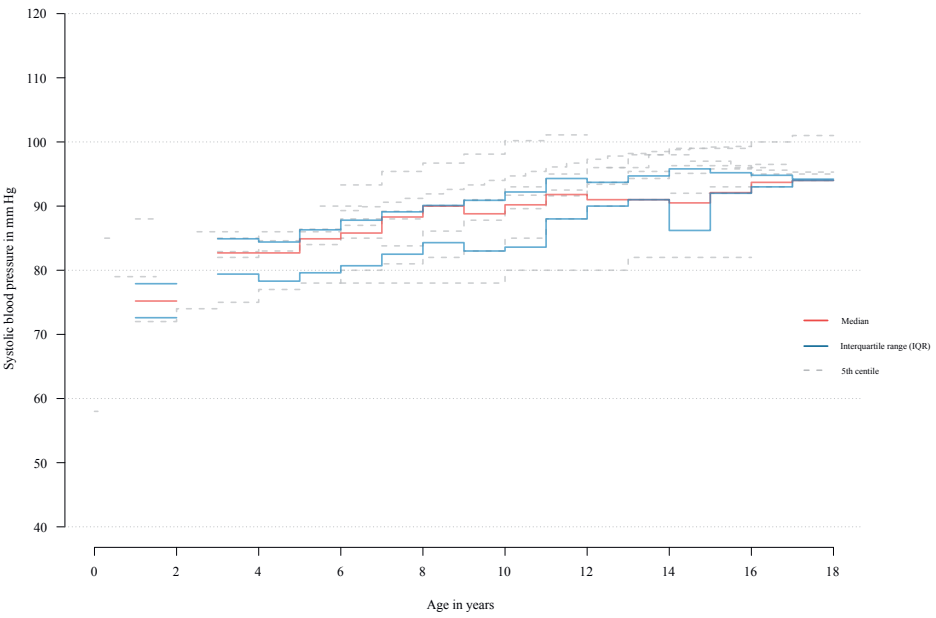
Additional file 2. Clinical definitions for hypotension and range of 5th centile of systolic blood pressure for girls according to age

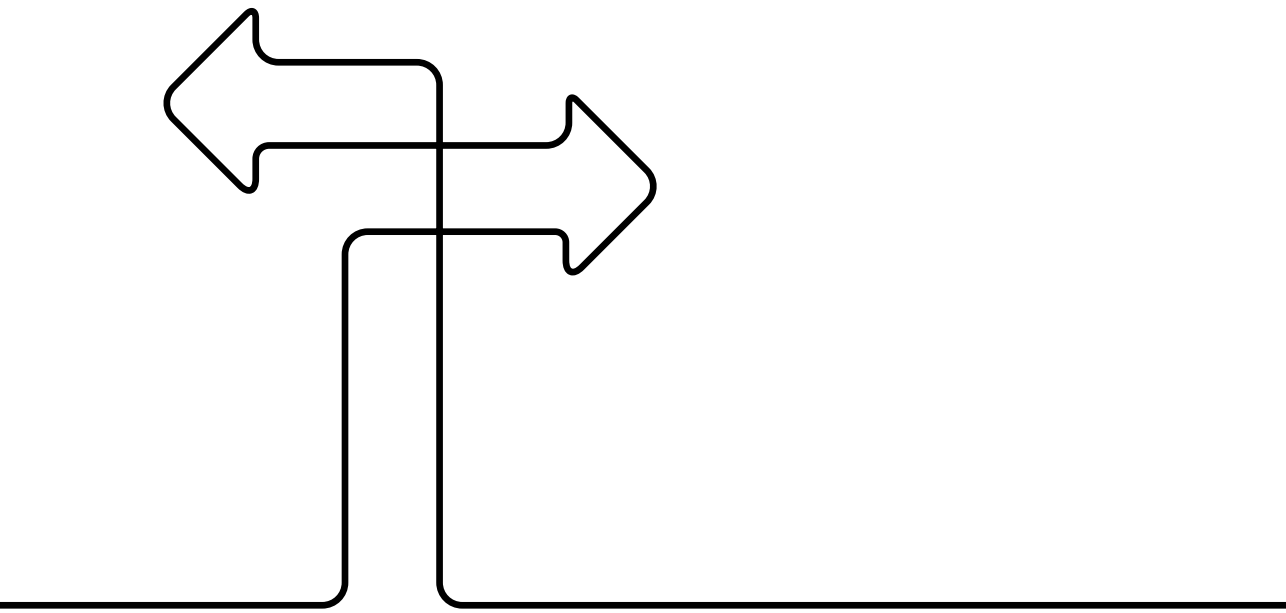


Additional file 3. 5th centile of systolic blood pressure and median (IQR) for boys



Additional file 4. 5th centile of systolic blood pressure and median (IQR) for girls





Chapter 10

Development and validation of a paediatric early warning score for use in the emergency department: a multicentre study

Joany M Zachariasse, Daan Nieboer, Ian K Maconochie, Frank J Smit, Claudio F Alves, Susanne Greber-Platzer, Maria N Tsolia, Ewout W Steyerberg, Paul Avillach, Johan van der Lei, Henriëtte A Moll

Lancet Child Adolesc Health. 2020 Aug;4(8):583-591

ABSTRACT

Background Paediatric Early Warning Scores (PEWSs) are being used increasingly in hospital wards to identify children at risk of clinical deterioration, but few scores exist that were designed for use in emergency care settings. To improve the prioritisation of children in the emergency department (ED), we developed and validated an ED-PEWS

Methods The TrIAGE project is a prospective European observational study based on electronic health record data collected between Jan 1, 2012, and Nov 1, 2015, from five diverse EDs in four European countries (Netherlands, the UK, Austria, and Portugal). This study included data from all consecutive ED visits of children under age 16 years. The main outcome measure was a three-category reference standard (high, intermediate, low urgency) that was developed as part of the TrIAGE project as a proxy for true patient urgency. The ED-PEWS was developed based on an ordinal logistic regression model, with cross-validation by setting. After completing the study, we fully externally validated the ED-PEWS in an independent cohort of febrile children from a different ED (Greece).

Findings Of 119,209 children, 2007 (1·7%) were of high urgency and 29,127 (24·4%) of intermediate urgency, according to our reference standard. We developed an ED-PEWS consisting of age and the predictors heart rate, respiratory rate, oxygen saturation, consciousness, capillary refill time, and work of breathing. The ED-PEWS showed a cross-validated c-statistic of 0·86 (95% prediction interval 0·82 to 0·89) for high urgency patients and 0·67 (0·61 to 0·73) for high-urgency or intermediate-urgency patients. A cutoff of score of at least 15 was useful for identifying high-urgency patients with a specificity of 0·90 (95% CI 0·87 to 0·92) while a cutoff score of less than 6 was useful for identifying low-urgency patients with a sensitivity of 0·83 (0·81 to 0·85).

Interpretation The proposed ED-PEWS can assist in identifying high-urgency and low-urgency patients in the ED, and improves prioritization compared with existing PEWSs.

INTRODUCTION

Worldwide, emergency departments (EDs) struggle with the continuously increasing demand for emergency care. With this increasing burden on EDs, concerns have been raised about the effect on the quality of these services for children.[3, 8, 50, 206] Recognising those children who require immediate attention amidst the large group of children who do not require urgent care is considered essential for ensuring patient safety, particularly in overcrowded EDs. Most EDs use a triage system to prioritise all visiting patients, including children. However, research shows that these systems still do not identify a substantial proportion of children with serious illness.[13, 34, 128]

Vital signs are considered an essential tool in the assessment of a patient's clinical condition. They are objective measures, do not require spoken language, and can be obtained relatively fast by trained health-care workers. The combination of multiple physiological measurements appears to be a promising tool to identify children with serious illness.[207, 208] Scoring systems based on physiological measurements, so-called Paediatric Early Warning Scores (PEWSs), have been developed to detect clinical deterioration in patients admitted to hospital by repeatedly measuring scores and observing trends over time.[121] In emergency settings, these PEWSs are increasingly being applied by health-care workers during the first assessment of paediatric patients, to aid in the recognition of seriously ill children or those at risk of deterioration.[209, 210] The same scores, originally developed for the inpatient setting, are now being used in EDs.[211-214] However, children admitted to hospital wards have already been identified as having some medical need, whereas the general ED population typically includes a large group of relatively well children with self-limiting conditions. Furthermore, the currently available PEWSs were all established by expert opinion, often without validation, and therefore their performance in emergency settings is unclear.[208, 209, 215] A new PEWS designed for use in the ED is needed that can accurately identify the small group of seriously ill children who require immediate care, but can also rule out serious illness in the large group of non-urgent patients. So far, no PEWS exist that is applicable to the paediatric ED population, based on real-world data, and validated for use in emergency care settings.

We aimed to develop and validate a PEWS, based on a large multinational cohort, to improve the prioritisation of children visiting the ED.

METHODS

Study design and participants

The study was embedded in the TrIAGE project, a prospective observational study aiming to improve the early recognition of seriously ill children in the ED. In the TrIAGE study, electronic health record data were included prospectively from consecutive, non-scheduled ED visits of children children under age 16 years from five diverse EDs in four European countries: Erasmus MC (Rotterdam, the Netherlands), Maasstad Hospital (Rotterdam, the Netherlands),St Mary's Hospital (London, UK), Hospital Fernando da Fonseca (Lisbon, Portugal), and Vienna General Hospital (Vienna, Austria). Enrolment varied by study site and took place during a period of 8-36 months between Jan 1, 2012, and Nov 1, 2015 (Supplement 1). The five participating study sites were diverse regarding type of hospital, number of ED visits, and complexity of the patient population (Supplement 1). All hospitals used an electronic hospital information system in which ED nurses routinely entered the collected clinical data. Triage was routinely performed with the Manchester Triage System (MTS)[14]. Data on patient characteristics, physiological parameters, and outcome were extracted automatically into a database, harmonised, and checked for quality. The study was approved by the medical ethical committees of the participating institutions. All waived the requirement for informed consent.

Predictors

PEWSs are scoring systems consisting of physiological parameters, typically with age-related cutoff values, that should be easy to calculate manually. To develop such a score, we built an ordinal regression model, and, based on this complex model, we derived a simple score, the ED-PEWS. We identified candidate predictors through a review of existing PEWSs.[131] Variables were selected if they were physiological measurements, regularly measured in the initial assessment in the ED. Four key physiological parameters were identified that were present in almost all published scores: heart rate, respiratory rate, oxygen saturation, and consciousness. Capillary refill time and work of breathing were considered as potential additional variables. Finally, temperature and pain score were selected because these influence other vital signs. Blood pressure, although important in adults, was not included because of its limited value when routinely done in the unselected population of children visiting the ED, and because it was not a standard measurement in the study sites.[172] At the participating EDs, physiological parameters were measured at discretion of the nurse, according to local practices (Supplement 1). All physiological measurements were explored using cross-tabulations, histograms and box plots. Values below the 0·01 or above the 0·99 percentile values were judged to be implausible and truncated.[216] Missing physiological measurements were imputed 25 times using the MICE algorithm in R (version 3.6.3; Supplement 2).[115] We assumed these items to be missing at random, which means that missingness can be fully accounted

for by other variables in the database[141], because we expected strong associations between patient factors (eg, patient characteristics, type of referral, presenting problem, and triage urgency) and setting factors (type of hospital, and month, day, and hour of presentation). The imputation model therefore included all predictors and outcome measures and additional descriptors of casemix, including patient age and sex, date and time of arrival, and triage characteristics. The imputation process resulted in 25 datasets on which statistical analysis were done and pooled for a final result.[141]

The variables heart rate, respiratory rate, oxygen saturation, and temperature were included as continuous to preserve maximum information. Linearity was tested with restricted cubic splines, using the rms library in R. The continuous variables temperature and oxygen saturation could be modelled adequately as a linear function. Heart rate and respiratory rate showed a non-linear relationship with the outcome and were modelled using restricted cubic splines with five knots. We assessed six relevant interactions, specifically the interaction of age, temperature, and pain, with both heart rate and respiratory rate. Because there was no significant interaction (data not shown), no interaction terms were added to the model.

Beside the physiological measurements, the variables age and setting were added as predictors

Age was converted into an ordinal variable with clinically relevant categories (<1 year, 1 year to <2 years, 2 years to <5 years, 5 years to <12 years, and ≥12 years), based on the cutoffs used in the advanced paediatric life support guidelines.[102]

Outcomes

As part of the TrIAGE project, a reference standard was developed, serving as a proxy for each child's true urgency (Supplement 3). This reference standard was developed based on the methodology of a previously published study, based on information from the entire ED visit.[69, 119] It consists of three categories: high, intermediate, and low urgency. These categories reflect the time before a patient should be seen by a physician. As secondary outcome measures, we used intensive care unit (ICU) and hospital admission immediately after the ED visit, because these are reference standards most commonly found in the literature. [13]

Model derivation and PEWS creation

First, an ordinal logistic regression model was derived on the full data set. We explored the assumptions underlying ordinal logistic regression and did not find any major violations. To avoid stepwise variable selection methods, we entered candidate predictors hierarchically in the model using three steps: 1) key variables only (heart rate, respiratory rate, oxygen saturation, and consciousness); 2) key variables plus possible additional variables (capillary refill time and work of breathing); and 3) key variables plus possible additional variables plus predictors affecting other variables (temperature and

pain score). All models included setting and patient age. Setting was added to take into account the multicentre nature of the study and adjust for confounding by study site. Age was added to adjust for the age-dependent normal values of several of the vital signs. The final model was selected based on performance according to the χ^2 statistic and expert opinion where we strive for the most parsimonious model.

We developed the ED-PEWS based on the full model using the nomogram function from the Hmisc package in R. This function uses the coefficients of the predictors to assign scores to the different variables. We compared performance of the ED-PEWS with the extensive model to see whether accuracy was reduced.

Performance assessment

An accurate model would discriminate between patients of high and low urgency. We quantified discrimination of the ED-PEWS with the c-statistic. The c-statistic ranges from 0·5 to 1, with a higher score indicating better discrimination. Because our reference standard has three ordinal categories, we report the c-statistic for two different cutoffs: the identification of high-urgency patients and the identification of high-urgency or intermediate-urgency patients. We present 95% prediction intervals, beside the 95% CIs, to show the effect of heterogeneity between settings on overall model performance.

Calibration refers to the level of agreement between predicted risks and observed outcome. We assessed calibration with a calibration plot, comparing the predicted risks with the observed proportions of high-urgency or intermediate urgency outcomes. The ideal slope of such a plot is 1, indicating perfect agreement between observed and predicted risks

Decision curve analysis is a method of evaluating the performance of prediction models, taking into account their clinical consequences.[109] In the ED, a new PEWS might improve the early identification of high-urgency patients, but might also lead to a large number of false positives that hamper ED workflow. Therefore, a threshold probability is required: how many false-positive patients does one accept to find one truly high-urgency patient? A decision curve plots the net benefit of the model, over a range of these threshold probabilities, and allows for comparison between different clinical alternatives. The model with the highest net benefit over a given threshold probability has the largest clinical value.

We used internal-external validation to assess performance of the ED-PEWS. We applied leave-one-out cross-validation by omission of data from each hospital in turn. Thus, we constructed the PEWS based on data from four hospitals, assessed performance on the fifth hospital, and repeated this process five times. For the c-statistic, we pooled the resulting five estimates with a random-effects model. Calibration plots and decision curves were created for each of the five datasets separately.

Diagnostic accuracy measures (sensitivity, specificity, and positive and negative likelihood ratios) were calculated for several of the score's cutoff points. These measures

were calculated separately for each of the hospitals, and pooled using the glmer function from the lme4 package in R.

We assessed the effect of multiple imputation on the development of the ED-PEWS in a sensitivity analysis in which we fitted a model on a complete case dataset and compared the model and its coefficients with the original model.

Value in practice and external validation

As an illustration, we compared the performance of the ED-PEWS with two existing PEWSs that have been applied previously in the ED setting.[131, 140, 160, 217] The Paediatric Advanced Warning Score by Egdell and colleagues is one of the few available scores developed specifically for use in the ED. This score was developed to identify children visiting the ED who were in need of urgent medical assessment as reflected by ICU admission after the ED visit.[217] The Bedside PEWS by Parshuram and colleagues was developed to identify clinical deterioration in hospitalised children, but has been the only PEWS so far to be evaluated rigorously in a multicentre randomised clinical trial.[140, 160] We had to adjust these scores based on the variables that were available in our database and thus had to exclude blood pressure (not available), and oxygen therapy (part of our reference standard), and we used different categories for the variables respiratory effort; capillary refill time; and the alert, verbal, pain, unresponsive scale. We assessed the performance of the two scores, maintained as a continuous variable, for each of the reference standards and calculated the pooled c-statistic.

To gain more understanding on the value of the ED-PEWS in routine care, we assessed its additional value above regular triage routinely used in the ED. Therefore, we assessed the significance of the ED-PEWS in a model adjusted for triage classification by the MTS (the routinely used triage system in all hospitals) with the reference classification as the outcome. Moreover, we calculated the c-statistic for the MTS alone and for the combination of MTS and ED-PEWS to show the increase in discrimination.

Finally, after completing the study, we fully externally validated the ED-PEWS in an independent cohort of febrile children from the P and A Kyriakou Children's Hospital (Athens, Greece). This hospital is one of the two large public children's hospitals in the greater Athens regions. It was selected because a large volume of children attend its ED, who are mostly of low urgency: a population that is common in western European countries (Supplement 1). Children were included during 2 random weeks each month between January, 2017, and April, 2018, if they presented with fever to the ED (temperature ≥38.0°C) or history of fever in the 72 h before the ED visit. All required items for the ED-PEWS and reference standard were available, except oral medication administered in the ED, as part of the intermediate-urgency reference classification.

Role of the funding source

The funders of the study had no role in the study design, data collection, data analysis,

data interpretation, writing of the report, or the decision to submit for publication. The corresponding author had full access to all the data in the study and had the final responsibility for the decision to submit for publication.

RESULTS

From Jan 1, 2012, to Nov 1, 2015, 119,209 children presented to the ED in one of the participating hospitals; exact study periods varied by hospital (Supplement 1). None of these ED visits were excluded. The median patient age was 4·4 years (IQR 1·7-9·5) and 54,836 children (46·0%) were girls. According to our reference standard, 2007 children (1·7%) were of high urgency (range across hospitals 0·6% to 7·2%) and 29,127 (24·4%) of intermediate urgency (range across hospitals 20·2% to 39·6%). After the ED visit, 11,754 children (9·9%) were admitted to hospital and 698 (0·6%) to ICU. Although most patients were considered of low urgency, the case-mix was diverse and differed among the different participating sites (Table 1). Distributions of the predictor variables are shown in the appendix (Supplement 4). In terms of the data available for each clinical variable, consciousness was reported most frequently (82·4%), while capillary refill time (48·4%) and respiratory rate (47·6%) were reported the least (Supplement 2).

All variables except temperature were significantly associated with the reference standard classification in the multivariable model (Table 2). The strongest predictors of patient urgency were decreased consciousness (odds ratio 8·55, 95% CI 6·68-11·0) and increased work of breathing (OR 7·07, 95% CI 6·66-7·50 for mild-to-moderate increase and 7·36, 6·21-8·71 for severe increase).

The most extensive ordinal regression model including all candidate predictors had the best performance. The model with the four key variables heart rate, respiratory rate, oxygen saturation and consciousness had a χ^2 statistic of 7,854 (based on the likelihood-ratio test). Adding the variables capillary refill time and work of breathing increased the χ^2 statistic to 11,896. Adding pain score and temperature further improved the model (χ^2 statistic 13,498). However, since the third step of model building only improved performance slightly, we decided to omit the variables temperature and pain score from the model.

Thus, we derived an ED-PEWS that included patient age and six predictors: heart rate, respiratory rate, oxygen saturation, consciousness, capillary refill time, and work of breathing (Figure 1).

Table 1. Demographics and baseline characteristics

	Erasmus MC (n=18,594)	Maasstad Hospital (n=10,584)	St Mary's Hospital (n=15,556)	Hospital Fernando da Fonseca (n=53,175)	General Hospital, Vienna (n=21,300)
Sex					
Male	10,774 (58)	6,004 (57)	8,677 (56)	27,685 (52)	11,233 (53)
Female	7,820 (42)	4,580 (43)	6,879 (44)	25,490 (44)	10,067 (47)
Age, years					
< 1	3,680 (20)	1,639 (15)	2,773 (18)	7,090 (13)	3,441 (16)
1 to < 12	11,855 (64)	6,556 (62)	10,742 (69)	38,447 (72)	15,239 (72)
≥ 12	3,059 (16)	2,389 (23)	2,041 (13)	7,638 (14)	2,520 (12)
MTS urgency					
Emergent or Very urgent	2,427 (13)	1,515 (14)	1,605 (10)	6,222 (12)	1,084 (6)
Urgent	8,745 (47)	5,110 (48)	3,961 (25)	10,951 (21)	3,851 (18)
Standard or Non-urgent	6,852 (37)	3,857 (36)	9,990 (64)	36,002 (68)	15,314 (72)
Unknown	570 (3)	102 (1)	0	0	1,051 (5)
Diagnostics					
Laboratory	6,234 (34)	2,102 (20)	1,558 (10)	6,990 (13)	7,841 (37)
Imaging	4,487 (24)	3,907 (37)	2,256 (15)	12,624 (24)	1,677 (8)
Medication					
Inhalation	762 (4)	612 (6)	1,073 (7)	5,223 (10)	951 (4)
Intravenous	2,182 (12)	1,360 (13)	703 (5)	4,009 (8)	989 (5)
Disposition					
ICU admission or death at ED	520 (3)	17 (0·2)	26 (0·2)	136 (0·3)	15 (0·1)
Hospital admission	3,801 (20)	2,463 (23)	1,599 (10)	2612 (5)	1,279 (6)
Discharge or other	14,273 (77)	8,104 (77)	13,931 (90)	50,427 (95)	20,006 (94)

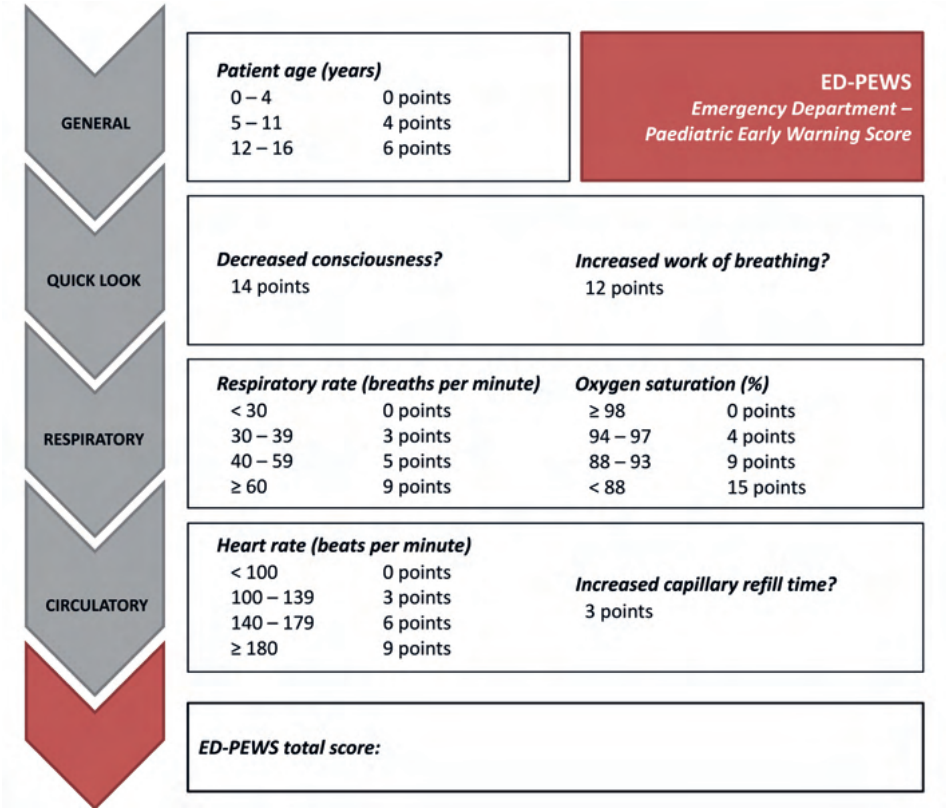
Data are numbers (%). Data shown are without imputation. ED=emergency department. ICU=intensive care unit. MTS=Manchester Triage System.

Table 2. Association between hospital and the predictor variables with the reference standard classification in the total emergency department population

	Outcome		Odds ratios (95% CI)	
	High urgency	**Intermediate urgency**	**Univariable**	**Multivariable**
Hospital				
Erasmus MC	1,335/18,594 (7·2)	5,940/18,594 (31·9)	*Reference*	*Reference*
Maasstad Hospital	129/10,584 (1·2)	4,193/10,584 (39·6)	0·97 (0·92-1·02)	0·77 (0·72-0·81)
St. Mary's Hospital	275/15,556 (1·8)	3,211/15,556 (20·6)	0·43 (0·41-0·45)	0·51 (0·48-0·54)
Hospital Fernando da Fonseca	1,163/53,174 (2·2)	10,717/53,174 (20·2)	0·43 (0·41-0·44)	0·49 (0·47-0·52)
General Hospital Vienna	127/21,300 (0·6)	4,212/21,300 (19·8)	0·37 (0·36-0·39)	0·45 (0·43-0·47)
Age, years				
<1	772/18,624 (4·1)	5,196/18,624 (27·9)	*Reference*	*Reference*
1 to <2	454/15,230 (3·0)	3,614/15,230 (23·7)	0·77 (0·73-0·81)	0·93 (0·88-0·99)
2 to <5	688/30,324 (2·3)	6,532/30,324 (21·5)	0·66 (0·63-0·68)	1·10 (1·04-1·16)
5 to <12	748/37,284 (2·0)	837/37,284 (22·5)	0·68 (0·65-0·70)	1·66 (1·56-1·78)
≥12	367/17,747 (2·1)	4,558/17,747 (25·7)	0·80 (0·76-0·83)	2·28 (2·10-2·47)
Heart rate (first vs third quartile)	-	-	1·62 (1·56-1·69)	1·68 (1·58-1·78)
Respiratory rate (first vs third quartile)	-	-	1·50 (1·44-1·57)	1·34 (1·26-1·42)
100- oxygen saturation	-	-	1·28 (1·27-1·29)	1·16 (1·15-1·17)
Consciousness				
Normal	2,545/117,671 (2·2)	27,617 /117,671 (23·5)	*Reference*	*Reference*
Decreased	485/1,539 (31·5)	657/1,539 (42·7)	13·59 (10·68-17·29)	8·55 (6·68-11·0)
Temperature (per 5°C)	-	-	4·09 (3·81-4·38)	1·03 (0·93-1·14)
Pain score	-	-	1·09 (1·08-1·10)	1·11 (1·10-1·12)
Work of breathing				
Normal	1,637/110,221 (1·5)	23,024/110,221 (20·9)	*Reference*	*Reference*
Mild-to-moderate increase	1,112/8,126 (13·8)	4,794/8,126 (59·0)	9·53 (9·03-10·06)	7·07 (6·66-7·50)
Severe increase	272/862 (31·6)	456/862 (52·9)	25·38 (21·84-29·49)	7·36 (6·21-8·71)
Capillary refill time				
Normal	2,820/117,569 (2·4)	27,638/117,569 (23·5)	*Reference*	*Reference*
Abnormal	210/1,641 (12·8)	636/1,641 (38·8)	3·42 (3·03-3·85)	1·70 (1·49-1·95)

Data are n/N (%), unless otherwise specified.

Figure 1. The ED-PEWS



ED-PEWS=Emergency Department Paediatric Early Warning Score.

The cross-validated c-statistic of the ED-PEWS was 0·86 (95% prediction interval 0·82 to 0·89) for the identification of high-urgency patients and 0·67 (0·61 to 0·73) for the identification of high-urgency or intermediate-urgency patients (Table 3). Across the different hospitals, the c-statistic ranged from 0·82 to 0·90 for high-urgency, and from 0·64 to 0·71 for high-urgency and intermediate-urgency patients (Supplement 5). Regarding our secondary outcome measures, the cross-validated c-statistic was 0·83 (95% prediction interval 0·77 to 0·89) for ICU admission, and 0·69 (0·64 to 0·73) for hospital admission (Supplement 5).

Discrimination of our model was better than two other commonly used PEWSs, and net benefit was higher for most of the clinically relevant threshold probabilities (Table 3, Supplement 6).

Table 3. Discrimination of the ED-PEWS and comparison with other PEWSs

	Reference standard: high urgency vs intermediate or low urgency			Reference standard: high or intermediate urgency vs low urgency		
	c-statistic	**95% CI**	**95% prediction interval**	**c-statistic**	**95% CI**	**95% prediction interval**
ED-PEWS	0·86	0·84 – 0·88	0·82 – 0·90	0·67	0·64 – 0·69	0·61 – 0·73
Parshuram PEWS*	0·82	0·79 – 0·84	0·76 – 0·87	0·64	0·61 – 0·66	0·57 – 0·70
Egdell PEWS**	0·81	0·80 – 0·83	0·78 – 0·84	0·63	0·61 – 0·65	0·59 – 0·68

	ICU admission			Hospital admission		
	c-statistic	**95% CI**	**95% prediction interval**	**c-statistic**	**95% CI**	**95% prediction interval**
ED-PEWS	0·83	0·79 – 0·87	0·77 – 0·89	0·69	0·67 – 0·71	0·64 – 0·73
Parshuram PEWS*	0·79	0·71 – 0·87	0·62 – 0·97	0·65	0·62 – 0·67	0·60 – 0·69
Egdell PEWS**	0·78	0·72 – 0·85	0·66 – 0·91	0·64	0·62 – 0·66	0·59 – 0·68

ED-PEWS=Emergency Department Paediatric Early Warning Score. PEWS=Paediatric Early Warning Score. ICU=intensive care unit.

* Modifications made to enable calculation of score using our data: exclude systemic blood pressure (not available); exclude oxygen therapy (part of our reference standard); categorize respiratory effort into three categories instead of four.

** Modifications made to enable calculation of score using our data: categorise capillary refill time into two categories instead of three; categorize alert, verbal, pain, unresponsive scale into three categories instead of four

We observed substantial heterogeneity in the predicted risk of high or low urgency between the different settings. The calibration plots suggest that the ED-PEWS underestimates the risk of high urgency in settings with a high proportion of high-urgency patients, and overestimates the risk in settings with a relatively low proportion of high-urgency patients (Supplement 7).

Using a cutoff score of at least 15 for high urgency placed 11·9% of patients (range across hospitals 6·9% to 15·9%) in the high-urgency category (Table 4). This gave a specificity of 0·90 (95% CI 0·87 to 0·92) and a positive likelihood ratio of 6·8 (95% CI 5·3 to 8·4). Using a cutoff score of less than 6 for low urgency, placed 27·9% of patients (range across hospitals 18·4% to 32·4%) in the low-urgency category, with a sensitivity of 0·83 (95% CI 0·81 to 0·85) and a negative likelihood ratio of 0·53 (95% CI 0·48 to 0·58). Different cutoff scores can be applied to improve sensitivity and specificity (Table 4).

Table 4. Diagnostic accuracy of various Emergency Department Paediatric Early Warning Score cutoff points

	Proportion of patients classified (range over hospitals)	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Cut-off score for high urgency					
≥30	1·6% (0·6%-2·7%)	0·27 (0·21-0·25)	0·99 (0·99-1·00)	33·6 (19·3-47·9)	0·73 (0·67-0·80)
≥25	3·9% (1·5%-5·8%)	0·42 (0·38-0·47)	0·97 (0·96-0·98)	16·1 (9·9-22·3)	0·59 (0·54-0·64)
≥20	7·7% (3·5%-10·3%)	0·58 (0·55-0·61)	0·94 (0·92-0·96)	9·7 (6·8-12·6)	0·45 (0·42-0·47)
≥15	11·9% (6·9%-15·9%)	0·68 (0·65-0·72)	0·90 (0·87-0·92)	6·8 (5·3-8·4)	0·35 (0·31-0·39)
≥10	28·9% (24·9%-37·3%)	0·81 (0·78-0·84)	0·73 (0·68-0·77)	3·0 (2·6-3·4)	0·26 (0·22-0·30)
Cut-off score for low urgency					
<6	28% (18%-32%)	0·83 (0·81-0·85)	0·32 (0·27-0·37)	1·2 (1·2-1·3)	0·53 (0·48-0·58)
<7	47% (39%-52%)	0·67 (0·65-0·70)	0·53 (0·49-0·56)	1·4 (1·3-1·5)	0·62 (0·57-0·67)
<8	59% (49%-65%)	0·57 (0·54-0·60)	0·66 (0·62-0·70)	1·7 (1·5-1·8)	0·65 (0·61-0·68)
<9	64% (54%-69%)	0·53 (0·50-0·56)	0·71 (0·67-0·74)	1·8 (1·6-2·0)	0·66 (0·62-0·70)
<10	71% (63%-75%)	0·46 (0·43-0·49)	0·78 (0·75-0·81)	2·1 (1·8-2·4)	0·69 (0·65-0·73)

The cutoff score for high urgency distinguishes between high-urgency vs intermediate-urgency and low-urgency patients, and the cutoff score for low urgency distinguishes between low-urgency vs high-urgency and intermediate-urgency patients.

In a model adjusted for triage classification, the ED-PEWS was significantly associated with the reference standard (Wald test; Z-statistic 64·5, p<0·0001) in the overall study population and in each of the individual hospitals (data not shown). The c-statistic for the high-urgency patients improved from 0·84 (95% prediction interval 0·76-0·91) for the MTS alone, to 0·90 (0·83-0·96) for the MTS in combination with the ED-PEWS. For the high-urgency and intermediate-urgency patients the c-statistic increased from 0·69 (95% prediction interval 0·65-0·73) to 0·73 (0·69-0·76).

In the fully external validation in the cohort of 4,542 febrile children, the c-statistic of the ED-PEWS for recognition of high urgency patients was 0·86 (95% CI 0·81 to 0·91) and 0·62 (0.60 to 0.64) for high-urgency and intermediate-urgency patients. The c-statistic for the secondary outcomes were 0·95 (95% CI 0·89 to 1·00) for ICU admission and 0·61 (0·58 to 0·63) for hospital admission. Specificity for the high-urgency cutoff was 0·95 (95% CI 0·90 to 1·00), compared with 0·90 (0·87 to 0·92) in the original cohort, and sensitivity for the low-urgency cutoff was 0·82 (95% CI 0·80 to 0·84) compared with 0·83 (0·81 to 0·85)

in the original cohort (Supplement 8). We repeated the model development process in a complete case dataset. Coefficients in the complete case analysis were largely similar and did not lead to changes in variable selection (Supplement 9). Moreover, the apparent performance in the development population was lower for the model developed in the original (imputed) database.

In the appendix (supplement 10), we present the results of applying the full ordinal regression model including all candidate predictors. This full model showed a cross-validated c-statistic of 0·86 (95% prediction interval 0·82 to 0·89) for the identification of high-urgency patients and 0·69 (0·63 to 0·74) for high- to-intermediate urgency patients.

DISCUSSION

In this large observational study in an unselected population of children presenting to the ED, a newly developed ED-PEWS was able to distinguish children with high and low urgency. To the best of our knowledge, this is the first validated PEWS to be developed specifically for use in the ED based on statistical modelling.

The ED-PEWS was developed based on a large dataset from five diverse EDs in different countries. We derived and validated our PEWS according to the conventions of traditional prediction models.[124, 216] For example, we handled missing data by multiple imputation, avoided stepwise selection methods, and maintained continuous physiological parameters as continuous. Thereby we aimed for an optimum accuracy within and outside the study population.

So far, no consensus exists regarding which PEWSs should be used in the ED. Many different PEWSs have been published, each consisting of different types and numbers of physiological parameters with diverse cutoff levels.[121, 131, 215, 218] Only one published PEWS used the discriminative ability of the candidate parameters in the predictor selection process, but combined the results with clinical judgement and used expert opinion to identify cutoff points for the different items.[140] Moreover, most PEWSs have been developed for use in hospitalized children.[121, 131] Such scores are likely to be of little value in the ED setting, because the baseline characteristics and risk profile of these children are very different. The new ED-PEWS outperforms two existing PEWSs when applied to the emergency care population and therefore appears to have additional value over previously published hospital-validated scores.

Our study has some limitations. Although ED staff were encouraged to report all required items as completely as possible, the recording of data and the measurement of physiological parameters was ultimately based on the discretion of the nurse. Therefore, we had to deal with missing data. The proportion of missing measurements was high,

most likely representing clinical practice, and in line with data from previous studies.[103, 134, 135] We used a multiple imputation approach to reduce bias by missing physiological measurements in the development of our model. The sensitivity analysis in a complete case dataset showed that the model remained largely the same. In addition, apparent performance in the development population was lower for the model developed in the original (imputed) database, thereby showing that our imputation process did not result in overestimation of performance.

Also, no gold standard exists that reflects patient urgency in the ED.[13, 29] Therefore, we developed a reference standard (high, intermediate, and low urgency), based on literature and expert opinion, as a proxy for true patient urgency. This reference standard consists of three outcome categories: high, intermediate and low urgency. Thereby, we aimed to reflect the prioritisation process in the ED, where a first step is to identify the high-urgency patients who require immediate attention, but a second step is to identify the low-urgency patients who can be allowed to safely wait for some time. We used hospital admission and ICU admission as secondary outcome measures; however, these measures do not fully reflect patient urgency and require dichotomization.

Additionally, although the normal values of heart rate and respiratory rate are related to age, the ED-PEWS does not include any age-specific cutoff values. We tested for several clinically relevant interactions, none of which were significant. Therefore, the addition of age-specific cutoff values would not improve our model and the items were only added as independent variables. This also improves the ease of use.

Finally, we did not include blood pressure as a potential predictor variable although hypotension is considered a late sign of deterioration and used for the diagnosis of shock. Blood pressure was not routinely done in the participating study sites. Furthermore, in a previous study, blood pressure was of little value in the unselected population of children in the ED,[172] and blood pressure is not recommended as an initial screening tool in the ED by several practice guidelines.[219, 220]

The newly developed ED-PEWS consists of a combination of physiological measurements that can be easily and objectively obtained in the emergency settings. Our results support its value in the prioritisation of children in the ED. Most EDs in high-income countries are visited by a lot of children who have mild or self-limiting conditions. Amidst those relatively well children, identification of the small group of children with serious illness or at risk of clinical deterioration is crucial. In our ED-PEWS, using a cut-off of at least 15, high-urgency patients can be identified with high specificity of 0.90 (95% CI 0·87 to 0·92). This high specificity is important, because these patients need to be seen by a physician immediately. Classifying too many low-urgency patients incorrectly as high urgency would increase the waiting time for the truly high-urgency patients and make the system less efficient. A cutoff of less than 6 can be used to rule out high-urgency and intermediate-urgency patients with a sensitivity of 0·83 (95% CI 0·81 to 0·85). The high sensitivity is important for this category,

to avoid false classification of high-urgency patients in the lowest urgency category, which may lead to seriously ill patients having to wait too long to be seen.

The ED-PEWS can be applied in diverse emergency care settings, although calibration suggests that the result might underestimate the risk of high urgency in settings with a high proportion of high-urgency patients, and overestimate the risk in settings with a low proportion of high-urgency patients. In the fully external validation in a cohort of febrile children with low urgency, performance was similar for the identification of high-urgency patients and somewhat poorer for the high-to-intermediate urgency patients. A limitation is that this cohort was from a single centre, with data on the specific subgroup of febrile children, and thus further exploration in different subgroups is required. Furthermore, we missed the variable oral medication as part of the intermediate-urgency reference standard category and therefore we might have underestimated the performance of the ED-PEWS for this cutoff. However, performance was also lower for the secondary outcome measure of hospital admission, which is part of the intermediate-urgency classification. Further studies are needed to explore heterogeneity in the performance of the ED-PEWS in specific subgroups of children.

The purpose of the ED-EWS is to identify high-urgency and low urgency patients, which is similar to the goal of a triage system. However, traditional triage systems are algorithms based on a wide variety of signs and symptoms. Moreover, they have a formal governance structure, undergo regular updates, and have standard implementation guidelines and training programmes available. Therefore, we do not propose that the ED-PEWS should replace a triage system. Rather, it should be used independently or as an adjunct to an existing triage system. The ED-PEWS could be used in clinical categories of children at high risk of undertriage. Future studies should focus on identifying subgroups of patients who would benefit most from additional triage with a PEWS.

A study done in the Netherlands reported that a third of hospitals use a PEWS in their ED, and that these 26 hospitals were using 20 different versions, with 18 different parameters in various combinations.[221] Consensus on which PEWS to use in the ED would enable comparison between EDs and facilitate future multicentre studies.

Although our ED-PEWS shows promising results, its performance indicates there is still room for improvement in the identification of high-urgency and low-urgency patients in the ED. For example, some children at risk of deterioration might present with physiological parameters that are still within the normal range. Further work should establish the value of other clinical predictors, including additional patient-related variables, nurses' or parental gut feeling, new biomarkers, and sequential vital sign measurements in the ED.

Finally, although the ED-PEWS is easy to calculate by hand, a slight improvement in performance can be achieved by using the full ordinal regression model. Ideally, hospitals should have the ED-PEWS and other decision rules implemented in their electronic medical records to facilitate its use by clinicians.

ACKNOWLEDGEMENTS

We gratefully acknowledge the emergency department nursing staff of the participating hospitals for their participation and collection of data. We thank Nienke Seiger, paediatrician (Erasmus MC-Sophia, Rotterdam), Rikke Jorgensen, formerly paediatric research nurse (St. Mary's Hospital, London) and Pinky Rose Espina, paediatric resident (General Hospital, Vienna) for their assistance with data collection for this study. We thank Dorine Borensztajn, paediatrician (Erasmus MC-Sophia, Rotterdam), for her contribution to the development of the TrIAGE database.

SUPPLEMENTARY INFORMATION

Available as online web appendix on the website of the Lancet Child & Adolescent Health:
- Supplement 1. Description of the different study sites
- Supplement 2. Missing values and multiple imputation
- Supplement 3. Reference standard validation
- Supplement 9. Complete case analyses
- Supplement 10. Performance of the full ordinal logistic regression model

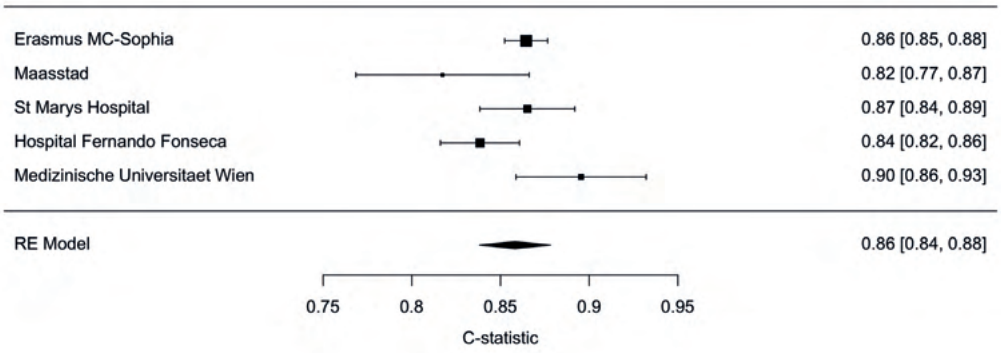
Supplement 4. Distribution of predictor variables and number of missings for each of the participating hospitals. Data shown without imputation.

	Erasmus MC (n=18,594)	Maasstad Hospital (n=10,584)	St Mary's Hospital (n=15,556)	Hospital Fernando da Fonseca (n=53,175)	General Hospital, Vienna (n=21,300)	Total missings (%) (n=119,209)
Age						
median (IQR)	43 (1·4-9·8)	5·7 (1·9-11·6)	3·9 (1·5-8·8)	4·7 (1·9-9·5)	3·9 (1·6-9·5)	0
Heart rate						
median (IQR)	120 (96-140)	130 (102-155)	118 (99-139)	116 (98-135)	115 (96-133)	50,898 (43)
Respiratory rate						
median (IQR)	28 (22-37)	36 (27-48)	26 (22-34)	24 (20-30)	24 (20-30)	62,496 (52)
Oxygen saturation						
median (IQR)	99 (97-100)	99 (97-100)	100 (98-100)	98 (97-100)	99 (99-99)	53,578 (45)
Consciousness, n(%)						
Normal	15,454 (83)	2,586 (24)	11,696 (75)	48,159 (91)	19,104 (90)	21,003 (18)
Decreased	899 (4·5)	101 (0·9)	54 (0·3)	76 (0·1)	77 (0·4)	
Temperature						
median (IQR)	37·2 (36·8-38·0)	37·2 (36·8-38·1)	36·9 (36·5-37·5)	36·3 (36·0-36·9)	37·0 (36·6-37·4)	26,822 (22)
Painscore						
median (IQR)	3 (1-4)	4 (2-5)	0 (0-2)	0 (0-3)	1 (0-3)	7,820 (7)
Work of breathing, n(%)*						
Normal	17,288 (93)	9,693 (92)	14,333 (92)	48,375 (91)	20,532 (96)	N/A*
Mild to moderate increase	1,128 (6)	572 (5)	1,073 (7)	4,594 (9)	759 (4)	
Severe increase	178 (1)	319 (3)	150 (1)	206 (0·4)	9 (<0·1)	
Capillary refill time, n(%)						
Normal	7,081 (38)	2,253 (21)	9,623 (62)	35,067 (66)	2,859 (13)	61,563 (52)
Abnormal	130 (0·7)	208 (2)	110 (0·7)	294 (0·6)	21 (0·1)	

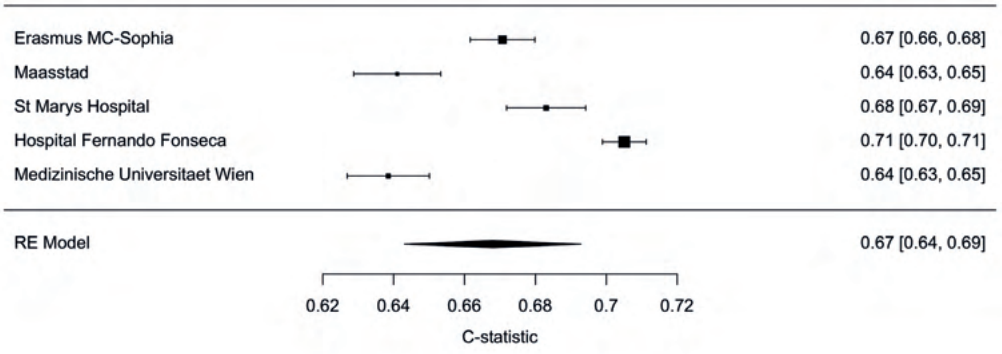
* No missings due to how variable was constructed

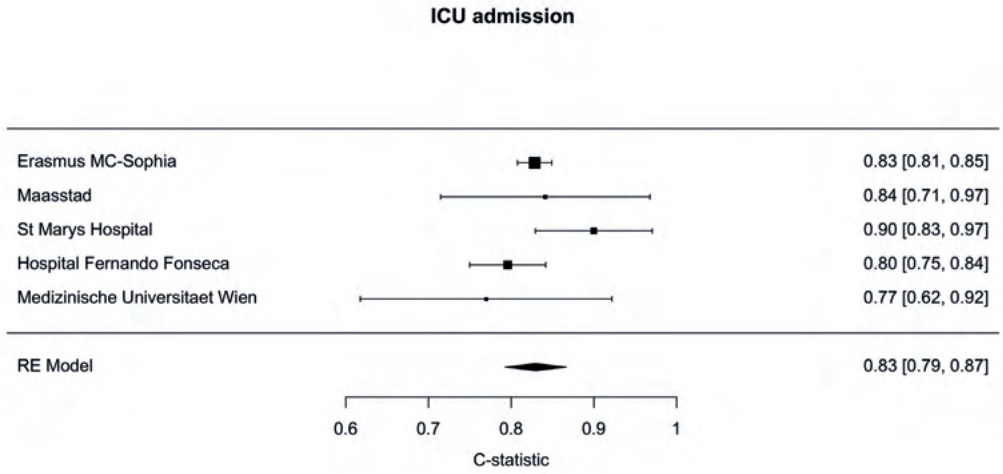
Supplement 5. Forest plots of the c-statistic in the leave-one-out cross-validation for each of the outcome measures

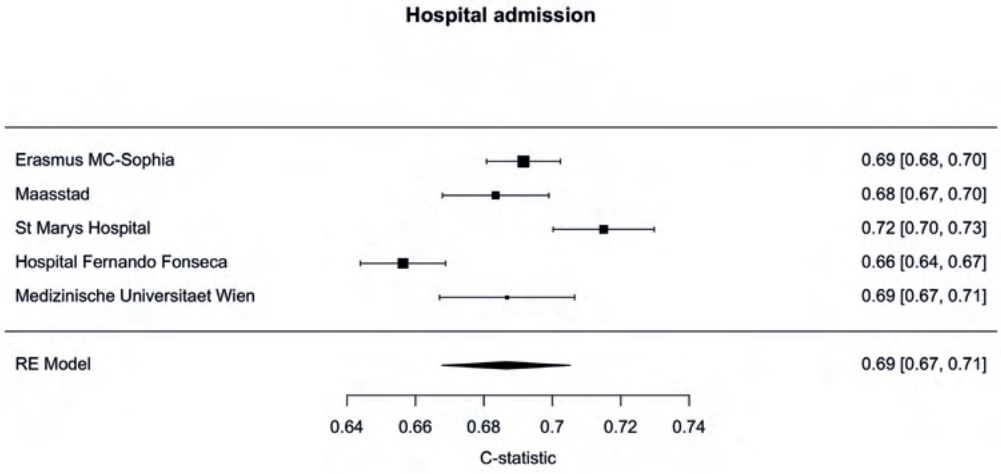
High urgency versus intermediate and low urgency reference standard classification



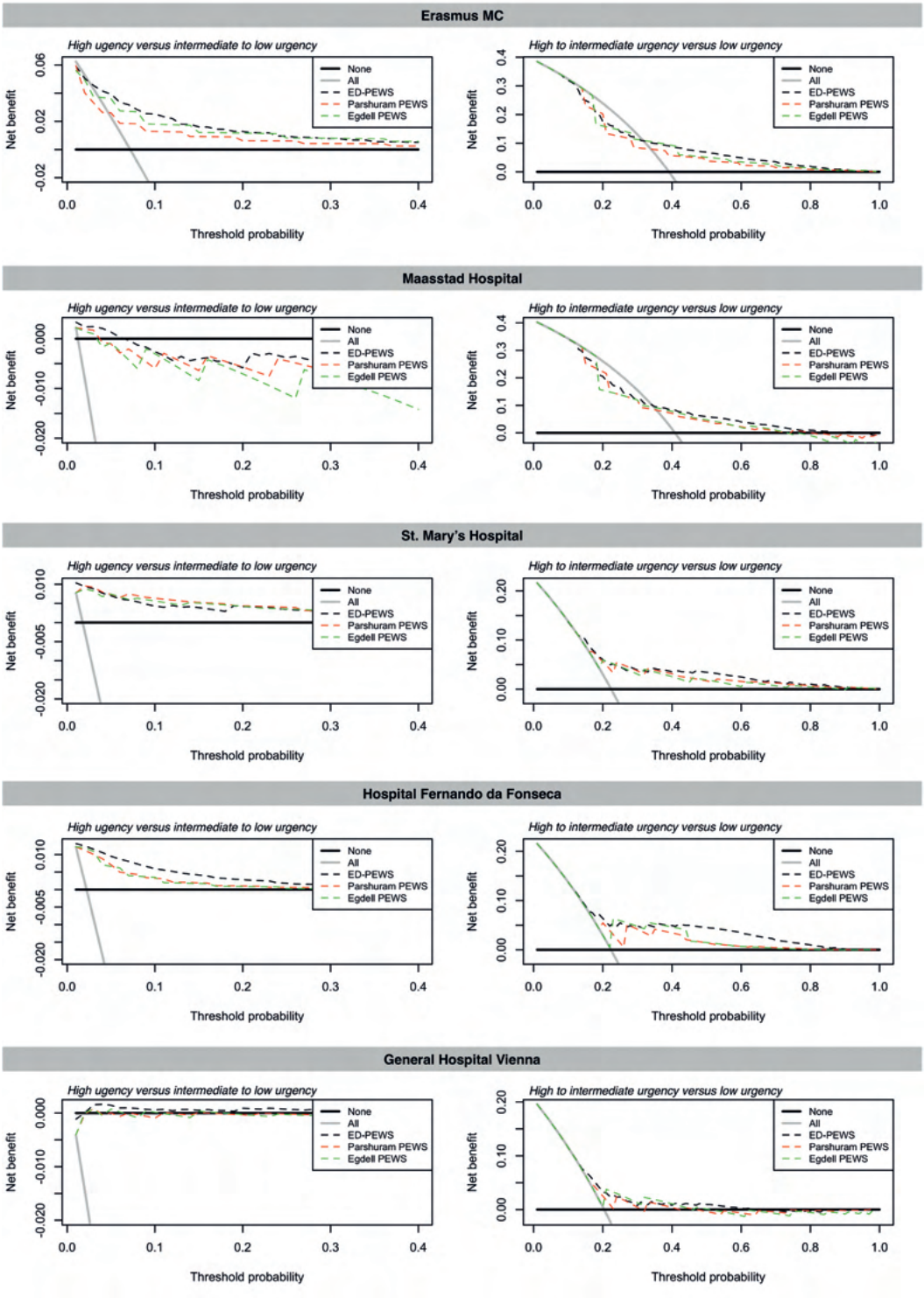
High and intermediate urgency versus low urgency reference standard classification



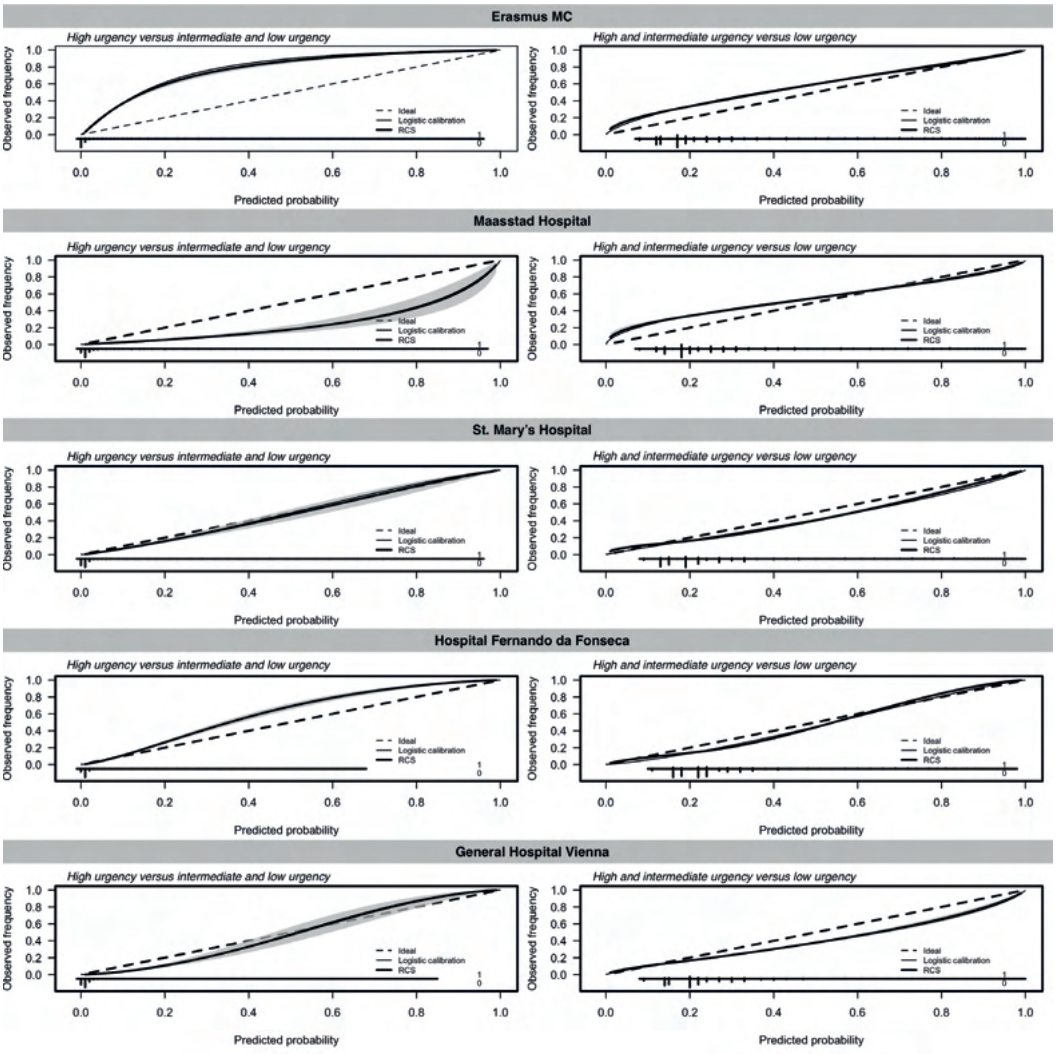

ICU admission
Erasmus MC-Sophia
Maasstad
St Marys Hospital
Hospital Fernando Fonseca
Medizinische Universitaet Wien
RE Model
0.83 [0.81, 0.85]
0.84 [0.71, 0.97]
0.90 [0.83, 0.97]
0.80 [0.75, 0.84]
0.77 [0.62, 0.92]
0.83 [0.79, 0.87]
0.6
0.7
0.8
0.9
1
C-statistic


Hospital admission
Erasmus MC-Sophia
Maasstad
St Marys Hospital
Hospital Fernando Fonseca
Medizinische Universitaet Wien
RE Model
0.69 [0.68, 0.70]
0.68 [0.67, 0.70]
0.72 [0.70, 0.73]
0.66 [0.64, 0.67]
0.69 [0.67, 0.71]
0.69 [0.67, 0.71]
0.64
0.66
0.68
0.7
0.72
0.74
C-statistic

Supplement 6. Decision curves of the ED-PEWS as compared to two existing PEWS stratified by hospital



Supplement 7. Calibration plots of the ED-PEWS stratified by hospital



Supplement 8. Fully external validation

Table 8.1 C-statistic of the ED-PEWS in the fully external validation

	Reference standard: High urgency versus intermediate and low urgency		Reference standard: High and intermediate urgency versus low urgency	
	C-statistic	**Confidence interval**	**C-statistic**	**Confidence interval**
ED-PEWS original performance	0·86	0·84-0·88	0·67	0·64-0·69
ED-PEWS performance in cohort of febrile children from P. and A. Kyriakou Children's Hospital	0·86	0·81-0·91	0·62*	0·60-0·64

	ICU admission		Hospital admission	
	C-statistic	**Confidence interval**	**C-statistic**	**Confidence interval**
ED-PEWS original performance	0·83	0·79-0·87	0·69	0·67-0·71
ED-PEWS performance in cohort of febrile children from P. and A. Kyriakou Children's Hospital	0·95	0·89-1·00	0·61	0·58-0·63

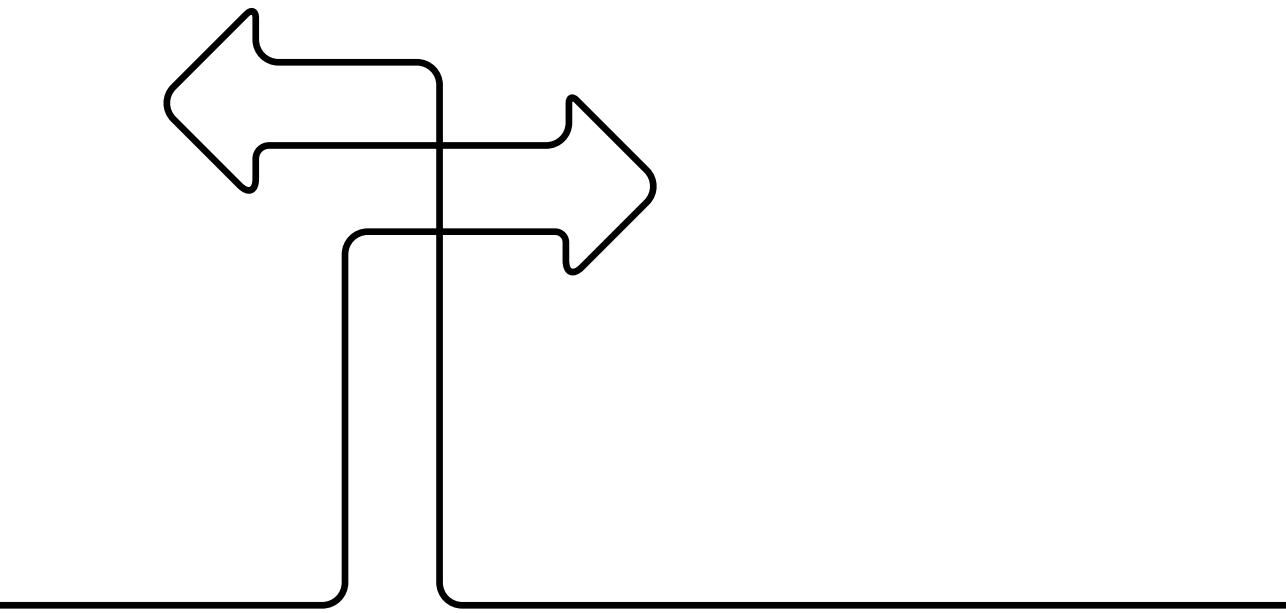
* Reference standard without oral medication in intermediate urgency category

Table 8.2 Diagnostic accuracy of the ED-PEWS in the fully external validation

Cut-off score for high urgency ≥15	**Proportion of patients classified as high urgent (range over hospitals)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Positive likelihood ratio (95% CI)**	**Negative likelihood ratio (95% CI)**
ED-PEWS original performance	11·9% (6·9%-15·9%)	0·68 (0·65-0·72)	0·90 (0·87-0·92)	6·8 (5·3-8·4)	0·35 (0·31-0·39)
P. and A. Kyriakou Children's Hospital	6%	0·66 (0·63-0·68)	0·95 (0·90-1·00)	12·8 (10·3-15·9)	0·36 (0·26-0·50)
Cut-off score for low urgency <6	Proportion of patients classified as low urgent (range over hospitals)	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
ED-PEWS original performance	28% (18%-32%)	0·83 (0·81-0·85)	0·32 (0·27-0·37)	1·2 (1·2-1·3)	0·53 (0·48-0·58)
P. and A. Kyriakou Children's Hospital	25%	0·82 (0·80-0·84)	0·27 (0·25-0·30)	1·1 (1·1-1·2)	0·65 (0·57-0·74)

PART IV

Discussion and summaries



Chapter 12

Discussion

DISCUSSION

In this thesis, we describe studies aiming to improve the first assessment of children presenting to the ED. In this discussion section, we will comment on several issues we encountered during this research. First we will address the gaps identified in prior research on triage systems. Subsequently, we will comment on how our findings contribute to the existing evidence. Finally, we will discuss future perspectives of research on the first assessment of children at the ED.

12.1 TRIAGE AT THE EMERGENCY DEPARTMENT: WHAT IS KNOWN AND WHAT ARE THE GAPS?

From the late 1990s, triage systems were implemented in EDs, to make sure patients, including children, were seen according to clinical need.[9] The triage systems were developed by the consensus of clinicians and nurses with experience in emergency medicine. Studies validating triage systems only became available after triage was already widely used in clinical practice. Still today, research on the performance of triage systems is scarce.[17]

The Manchester Triage System (MTS) is the most widely used triage system in Europe.[14] Previous studies have concluded that the MTS has a moderate validity in paediatric emergency care. It errs on the safe side with much more overtriage than undertriage.[31] In one study in children, modifications were proposed to reduce the amount of overtriage and improve the efficacy of the system.[35] Further efforts were aimed at identifying groups of risk for undertriage. It was found that young children, children with nonspecific medical problems, as well as children with comorbidity were particularly at risk. Moreover, it was observed that the majority of patients who were undertriaged had abnormal vital signs.[23]

Addressing the challenges in triage research

Beside the paucity of research on triage systems, available studies are often single centre, suffer from methodological limitations, and use a wide range of sometimes inadequate reference standards. Several challenges in research on triage systems need to be addressed, in order to provide high quality evidence that can have a direct clinical impact.

First of all, to study triage systems large numbers of patients are needed. Triage systems are usually complex algorithms. For example: the Manchester Triage System consists of more than 50 flowcharts and more than 180 distinct discriminators. The triage system is applied to the unselected ED-population, including children, adults and elderly, trauma and medical patients, and used for complaints from any possible organ system.

The proportion of high urgency patients in the ED is small and the presenting conditions diverse. To explore the validity of triage for certain subgroups of patients and have enough high urgency cases available, datasets should be large. Moreover, emergency departments within and between countries can be very heterogeneous regarding patient and hospital related factors (e.g. ED size, case-mix). Thus, to understand the performance of triage systems, multicentre and multinational studies are important.

Secondly, a single outcome measure that is reliable and accurate as a prognostic measure for urgency for the various conditions treated in the ED is lacking. Consequently, many available studies use inaccurate (e.g. interdependent with triage such as waiting time) and setting-dependent (e.g. costs) outcome measures. Intensive care unit (ICU) and hospital admission are two commonly used reference standards that are practical to use, but have important limitations. These are not true markers of priority, but rather of disease outcome (e.g. many conditions requiring hospital admission do not need immediate attention in the ED), and unlike triage systems they are dichotomous. Previously, three reference standards have been proposed as suitable for the evaluation of triage systems: an independent multi-level reference standard for urgency, immediate lifesaving interventions and expert opinion. Still, because published studies are based on a variety of reference standards, available research is difficult to compare and interpret.

Finally, triage systems are generally ordinal systems, classifying patients in 5 urgency categories. These categories correspond to the time a patient can safely wait before being seen by a physician. It is evident that for a triage system to be considered safe, it should accurately recognize the high urgency patients to prevent morbidity or even mortality. However, the ED population in many countries consist of a large population of low urgency, and a small population of high urgency patients. It is almost as important to accurately classify the group of low urgency patients. In case too many low urgency patients are falsely classified as high urgent, the waiting times for the truly high urgent patients increase and the system becomes inefficient. The majority of reference standards and performance measures used in published studies are only suitable for dichotomous analysis. To study the ordinal nature of triage systems, an ordinal reference standard and ordinal performance measures are needed.

The studies described in this thesis aim to provide knowledge on the performance of the MTS. Additionally, this thesis aims to improve triage by modifying the existing MTS and to study novel predictors and tools, such as early warning scores.

12.2 NEW INSIGHTS IN THE FIRST ASSESSMENT OF CHILDREN AT THE ED: WHAT THIS THESIS ADDS TO THE EXISTING KNOWLEDGE

Understanding the performance of triage systems

We conducted two studies on the performance of the MTS in large patient cohorts, with the objective to gain more insight in the MTS' validity, and identify patient subgroups at risk of undertriage. In both studies, one single centre study in children and one study in three hospitals including all age categories, specificity of the MTS was good, but sensitivity was moderate. Specific risk factors for undertriage were identified and included young age in children and old age in adults, medical presenting problem, and comorbidity.

In a systematic review summarizing the available evidence on the performance of triage systems, we found that performance of the MTS was similar to two other globally used triage systems (Canadian Triage and Acuity Scale and Emergency Severity Index). Nonetheless, a large heterogeneity in performance between different studies was observed, mostly due to differences in study design, the included population, performance measures and reference standards used. The results of these studies further emphasize that improvement of the MTS is needed, with a particular focus on the reduction of undertriage for the most vulnerable patients. The observed heterogeneity underscores the need for large and multinational studies, with an accurate and multilevel reference standard.

Vital signs in the first assessment of children in the ED

Physiological measurements have been shown important markers of disease severity in EDs and hospital wards, alone or in combination in early warning scores.[20, 121, 139] In a single centre study, we assessed the diagnostic value of blood pressure, an important vital sign in adults, for the recognition of serious illness in children. In this study, low blood pressure had low sensitivity to detect serious illness in the unselected group of paediatric ED visits. Therefore routine blood pressure measurement in the ED does not appear useful in the first assessment of children. This study was extended with a systematic review to identify population-based reference values for low blood pressure in healthy children and to compare these with cut-offs for hypotension defined by existing paediatric guidelines. We discovered that clinical guidelines show large variability and low to moderate agreement with population-based lower centiles. Remarkably, most clinical guidelines were based on expert opinion or on data from healthy children, and it is unclear how well these discriminate between healthy and diseased children in a clinical setting.

Clinical reference values such as the APLS guidelines or cut-off values from early warning scores are very heterogeneous and there is no consensus which values should be used in the ED. Moreover, it has been shown that these clinical reference values and

measurements from healthy children are partly overlapping.[133] Thus, there is a need for vital signs' reference ranges that are accurate for and applicable to children, and optimised for the purpose of triage at the ED.

Designing the TrIAGE project

In an attempt to further improve the triage of children, the TrIAGE (Triage Improvements Across General Emergency departments) project was devised. This project aimed to optimize the MTS for the triage of children at the ED through a large multicentre prospective observational study in different European emergency care settings. Building on prior experience, the project used a novel approach to overcome several of the previously discussed gaps in triage research. This included the use of routinely collected electronic health record (EHR) data to obtain a large sample size, the careful consideration of performance measures to assess modifications of triage systems, the definition of a three-category reference standard, and the combination of clinical knowledge and methodological strategies to identify modifications for triage systems.

The use of routine data from electronic health records

Since the publication of the first study on the MTS, electronic records have replaced paper files in the vast majority of European emergency departments. Using routinely collected clinical data for research purposes provides major opportunities. It enables the extraction of large quantities of data without time-consuming and expensive patient recruitment and data collection.

In the TrIAGE project, the inclusion of a large cohort was needed to enable the derivation of modifications for children at risk of misclassification, particularly in relatively rare subgroups. Furthermore, by including different clinical settings in several European countries, generalizability of the findings would be increased and heterogeneity between settings could be explored.

EHR data may lead to bias and misclassification if not used cautiously. The data in the TrIAGE project were carefully collected and checked by researchers with experience in paediatric emergency medicine. Site visits and checklist including each of the variables ensured accurate variable definitions and consistency among the settings. After data extraction, extensive quality checks were performed to further diminish the risk of errors and remove, for example, typing mistakes.

Methodological considerations

Most commonly used performance measures are only suitable for dichotomous outcomes while most emergency department triage systems classify patients into five ordinal categories. Recognizing that triage systems have multiple levels is important, because the systems are not only intended to identify the high urgency patients, but also distinguish the intermediate and low urgency patients. Moreover, different directions

of misclassification by a triage system have different clinical consequences. Safety of a triage systems is crucial, and thus reducing the amount of undertriage is paramount. On the other hand, any modification that improves the identification of high urgency patients will inadvertently result in additional false positive patients. Large increases in overtriage will decrease efficiency and applicability of the triage system and thus it is important that the trade-off between decrease in overtriage and increase in undertriage is conveyed by a performance measure.

To address the question what performance measures should be used to compare triage systems, we evaluated the properties of several performance measures in a case study and appraised each measure's strengths and weaknesses. Based on this appraisal, we proposed that as a minimum, decision curves and diagnostic accuracy measures should be used in a dichotomized analysis, and the triage-weighted kappa and Nagelkerke's R^2 in an ordinal approach. Although these are the most optimal performance measures available, they still have limitations including their dichotomous nature (decision curve and diagnostic accuracy measures), inability to take into account the weights of different types of misclassification (Nagelkerke's R^2), not being a proper scoring rule (diagnostic accuracy measures, triage-weighted kappa) and difficulties in interpreting differences in the performance measure (Nagelkerke's R^2 and triage-weighted kappa). Additionally, during our further studies the triage-weighted kappa appeared impractical to use, because the measure was relatively insensitive to changes, making it impossible to compare differences in the evaluation of triage modifications. Thus, there is still a need for performance measures in the evaluation of triage systems with multiple, ordinal, levels.

Defining a reference standard

We previously proposed an independent five-level reference standard as the most optimal reference standard available, but could not use this reference standard for the current study because it includes vital signs. Vital signs are used as predictors in the TrIAGE project and thus would result in interdependency. Moreover, vital signs are incorporated in other triage systems as well, making it less suitable for comparison between triage systems.

In practice, patient numbers in the highest and lowest triage levels are very low and thus in most studies the highest and lowest urgency categories are combined into three levels. Therefore, we developed a new three-level reference standard, based on an exhaustive literature review and a panel discussion of the expert members of the TrIAGE research group. Beside that items used in the triage system could not be included in the reference standard, reference standard items also had to be feasible (e.g. determining urgency by expert opinion of all patients in a large cohort study was considered too time-consuming), and independent of setting (e.g. costs are difficult to compare across countries).

Our approach resulted in a reference standard that classifies patients into three urgency levels and consists of multiple items related to patient urgency (immediate lifesaving interventions, resources used, disposition) as the best possible approximation of patient urgency. The multilevel nature enables assessment of discrimination between high and low urgency patients. Although data on the reference standard classification may not be as straightforward to collect as compared with single outcome measures such as hospital admission, the individual items are all routinely documented in most emergency departments. Hence, we encourage researchers studying triage systems to use this reference standard in future studies.

Combining clinical and methodological strategies to improve triage

To make use of available clinical evidence and optimize the triage for children, the TriAGE project combines clinical knowledge and methodological strategies (Figure 1).

All currently used triage systems were originally developed by expert opinion, and were validated only after implementation in practice. Ideally, however, modifications of triage systems, or new tools in the triage assessment, build upon available clinical evidence found in the literature, or after the analysis of real-world data extracted from EHRs.

In the evaluation of triage systems, it is crucial that accurate performance measures are used, and because there is no single optimal performance measure, multiple performance measures are required. It is known that a large heterogeneity in triage systems' performance across different settings exist. Therefore, multicentre studies are needed to estimate performance in different types of emergency care settings. Validation of a model is an important part of the model building process. It aims to assess a model's performance in different settings than where the model was derived. Leave-one-out cross-validation by setting is an example of an efficient validation strategy. In this approach, a model is derived in all settings but one, and validated in the setting that was left out of the model development, repeating this procedure so each of the settings is left out once. This method is efficient because it uses all available information, avoids wasting data by random splitting data in train and test sets, and illustrates heterogeneity in performance. Validation after implementation in practice is important to establish the value of a model in other settings, for example in low-resource settings and in different subgroups of patients.

Figure 1. Traditional approach to develop or modify triage systems compared with the novel approach applied in the TrIAGE study

Traditional approach	Novel approach
Developed as solution to clinical need	Developed as solution to clinical need or based on the evaluation of an existing approach
Selection of predictors based on expert opinion	Augmenting expert opinion with systematic literature review
Cut-offs from predictors based on normal values from healthy children or on available (expert-based) guidelines	Cut-offs from predictors based on statistical modelling using real-world clinical data from electronic health records
Performance measures mostly describe associations	Performance measures take into account multiple levels of triage urgency and the relative weights of over- and undertriage Multiple performance measures are used
Studies from a single centre, or multiple centres reported separately	Leave-one out cross-validation by centre with pooling of results to take heterogeneity into account
Validation *after* implementation in practice	Validation *before and after* implementation in practice in an ongoing process

Improving the MTS with vital signs

In the first study embedded in the TrIAGE project, we explored whether adding heart rate, respiratory rate and capillary refill time as discriminators to the MTS could improve the triage of children. Instead of using pre-specified cut-off values from clinical guidelines, we determined the optimal cut-off for each vital sign in the triage setting. Furthermore, we grouped all MTS flowcharts into nine clinical presentations (Cardiac, Dermatological, Ear Nose Throat, Gastrointestinal, Neurologic or Psychiatric, Respiratory, Trauma or Muscular, General malaise, Uro- or gynaecological and Other) and explored in which set of MTS flowcharts the new discriminators would provide additional benefit. As a result of the study, we propose to add six vital signs-based discriminators with optimal cut-off values to specific MTS flowcharts: "Very abnormal respiratory rate <1 year", "Very abnormal respiratory rate ≥1 year", "Abnormal heart rate <1 year", "Abnormal heartrate ≥1 year", "Abnormal respiratory rate <1 year" and "Abnormal respiratory rate ≥1 year". Adding these discriminators leads to a small but relevant increase in performance of the MTS.

A paediatric early warning score for use in the ED

In a second study embedded in the TrIAGE project, we developed and validated the Emergency Department - Paediatric Early Warning Score, shortened as ED-PEWS. To the best of our knowledge, this is the first PEWS specifically designed for use in the ED, based on statistical methods, that is able to distinguish high, intermediate and low urgency patients. It can be used independently or as an adjunct to formal triage systems. In the latter scenario, the ED-PEWS could be applied in patient subgroups at high risk of misclassification. Therefore, in a final study, we investigated the value of the ED-PEWS when applied in children with comorbidity, a vulnerable subgroup of patients at risk of undertriage. In this study in three European hospitals, the ED-PEWS had a similar

performance in children with comorbidity compared to previously healthy children. Adding the ED-PEWS to the existing triage system for children with comorbidity improved triage, except in the setting with relatively few high urgency patients.

Strengths and limitations of the TrIAGE project

Data collected in the TrIAGE project provided a unique opportunity to study and improve the triage of children. The large number of included children enabled the evaluation of specific subgroups and modifications targeted at the relatively rare high urgency cases. Because the project was conducted in five diverse EDs in four European countries we were able observe heterogeneity between settings. Because we used routinely collected EHR data, the project did not pose an additional burden on the ED staff. Moreover, by including all attending children risk of selection bias was low. The routinely collected data represent "real world" clinical practice. Finally, data were carefully collected and checked to ensure quality.

Despites these strengths, some limitations need to be discussed. First, although ED staff was encouraged to report all required items as completely as possible, the recording of data and the measurement of physiological parameters was ultimately based on the discretion of the nurse. Therefore, we had to deal with missing data, most importantly missing vital signs. We assumed that these missing physiological parameters were "missing at random", meaning that missingness can be fully accounted for by other variables in the database. We observed strong associations between the availability of the items and patient- and setting-related factors, supporting this assumption. We used a multiple imputation approach to deal with the missing variables, a common method to reduce bias by missing physiological measurements.

Second, although during the TrIAGE project a substantial number of variables were collected, we did not have detailed information on all aspects of the ED visit. For example, detailed information on presenting symptoms, test results, and working diagnosis at the end of the ED visit were not available. Additional data could have contributed to greater understanding of the determinants of patient urgency and could have contributed to additional predictors or specific groups at risk.

Finally, five EDs participated in the study which allowed us to take into account the heterogeneity between different settings. To fully explore the influence of ED characteristics, such as patient volume, case mix, or national health system, and triage characteristics, more settings are needed.

12.3 FUTURE PERSPECTIVES

Leveraging the value of electronic health record (EHR) data

Routinely collected EHR data are increasingly used for research purposes.[231, 232] Secondary use of these data presents many opportunities, but also many challenges.[233]

During the TrIAGE project, we encountered several practical difficulties extracting the data, such as limitations in data-extraction possibilities in EHRs, and challenges to customize the systems for research purposes. An additional challenge was the lack of standardization. One obvious factor in a European research project contributing to the difficulties in combining datasets is the use of different languages. Prior to the TrIAGE project, we discussed with all participating hospitals the availability of the variables and the quality of the data based on a quality control checklist to ensure the minimal requirements for consistency. During the project, a lot of time was spent on harmonizing the databases. For example, some hospitals recorded consciousness by the Alert, Verbal, Pain, Unconsciousness scale (AVPU), while others used the Glasgow Coma Scale (GCS), and others had their own scale. Each hospital used the MTS, but translated in their own language. Some hospitals already had modifications implemented based on the 3rd edition, while some hospitals had made their own adjustments. In addition, a major limitation in the use of large databases is the uncertainty about data quality. Even with the most careful checks before the study started, we encountered several issues during initial quality checks that would not have been discovered if data had not been checked in an early stage. Therefore, we propose that a prerequisite in the use of large datasets from different and possibly heterogeneous resources is that the data is carefully checked by individuals with expertise on the subject. Subject knowledge is crucial to ensure that findings make sense from the clinical point of view, resulting in further exploration of implausible values. The advantage of multicentre studies is that marked differences between settings may suggest potential errors.

Finally, beside technical and practical issues, ethical issues exist as well. Since the introduction of the European Union General Data Protection Regulation in 2018 (in Dutch: Algemene Verordering Gegevensbescherming), information from individuals is strongly protected. Personal data that directly, or indirectly in combination with other data sources, identifies an individual requires patients' consent, unless certain exemptions apply. These regulations may make the secondary use of EHR data more difficult.

Still, the use of EHR data offers great advantages. Through the use of EHRs, the TrIAGE project was able to obtain patient data from more than 100,000 ED visits. Obtaining such a large dataset would have been practically impossible if the data would have to be retrieved manually from the patient files.

Beside the large potential sample size, EHRs contain large amounts of detailed information. The potential to use these data may become larger, when information from

unstructured text becomes available through natural language processing techniques[234] or data can be linked to other data sources (GP files, biorepositories).[235] Finally, as EHRs consist of data that are collected for clinical purposes in near-real time, they offer the possibility to directly study the effect of changes in practice. For example, based on variables extracted from EHR data, a prediction rule can be developed, and its effect on outcome can be evaluated gathering EHR data after the implementation.

Future research projects should address the development of a large, nationwide database with data from ED visits. One example of such a large database is the publicly available National Hospital Ambulatory Medical Care Survey (NHAMCS), providing data on a national sample of visits to USA emergency departments.[236] In the Netherlands, accidents and injuries are reported in a representative sample of Dutch EDs in the Letsel Informatie Systeem (LIS).[237] However, no such database exists for other types of ED visits in the Netherlands yet. The development of such a database would be a great source for further research to improve the safety and efficacy of emergency care.

Combining clinical strategies and statistical modelling

All commonly used triage systems were developed by consensus from groups of experts. The same holds for all currently published PEWS for use in hospital and ED settings. Even several published vital signs reference ranges included in guidelines and textbooks are based on expert opinion. Our approach to derive a PEWS by regression modelling was therefore novel and the first published to date.

Expert knowledge from experienced physicians and nurses has been proven a valuable tool in the identification of predictors of patient urgency and ascertaining the range of physiological values that can be considered normal. However, when combining multiple variables or exploring interactions between different predictors, statistical modelling can be expected to be superior. In our study, the ED-PEWS outperformed two existing scores that were based on expert opinion, supporting the additional value of scores based on statistical modelling.

Existing vital signs reference ranges are heterogeneous and there is no consensus what the optimal cut-off value is. Future studies are needed to provide evidence-based recommendations. Moreover, as more data on physiological parameters becomes available, more complex modelling techniques are required to fully use the predictive potential of vital signs, such as models including repeated measurements and models based on monitor data.

Through the availability of large datasets and the advances in computer power, more advanced techniques such as machine learning and artificial intelligence have been introduced in clinical medicine and hold large promise to improve health care delivery. In the area of emergency medicine, there have been several published attempts to develop

a triage system by machine learning techniques.[238-242] To the best of our knowledge, these models have not been implemented in clinical practice. All of these studies had multiple important limitations such as the use of the area under the curve as the sole performance metric, the use of suboptimal outcome measures, a single centre design, the lack of external validation and/or the lack of an implementation study. In the wider field of medicine, the performance of machine learning algorithms, and its superiority over regression-based models is still under discussion.[243] [244, 245] Therefore, machine learning models should be welcomed with caution. Given the paucity of triage tools based on statistical modelling, focus should be on statistical models that are developed with sound methods, with predictors that make sense from the clinical point of view, that are easily applicable in ED practice and demonstrate a performance that is carefully assessed in well-designed studies.

Innovations in the first assessment of children

Although the MTS still appears in print, it has been implemented as an automated tool in the EHRs of most emergency departments. This clearly reflects its clinical value, because very few other clinical support systems are being implemented in EHRs on a large scale. The lack of integration in EHRs of, for example, PEWS or scores predicting hospital admission are important barriers in their usage as it limits their applicability and user-friendliness. As an example from the inpatient setting, a recent multicentre implementation study assessing whether use of the Bedside Paediatric Early Warning System (BedsidePEWS) could decrease mortality, required the nurse to obtain and manually record vital signs data on paper sheets, calculate the PEWS by hand, and look-up the score-matched recommendations.[160]

Thus, the next step in the implementation of PEWS should be their implementation in EHRs, and the possibility to provide automatic generation of practice recommendations for triage nurses. In a subsequent and more advanced scenario, PEWS could be developed that are based on monitor data and do not require nurses to record individual vital signs in patients' records. These models can also incorporate more advance characteristics of physiologic measurements (such as heart rate variability) and incorporate trends over time (such as decreasing or increasing values). Automated deterioration detection based on monitor data has been studied in the intensive care unit[246, 247], although with varying results. Their value in the ED population, a much more diverse population with on average lower illness severity, is unclear.

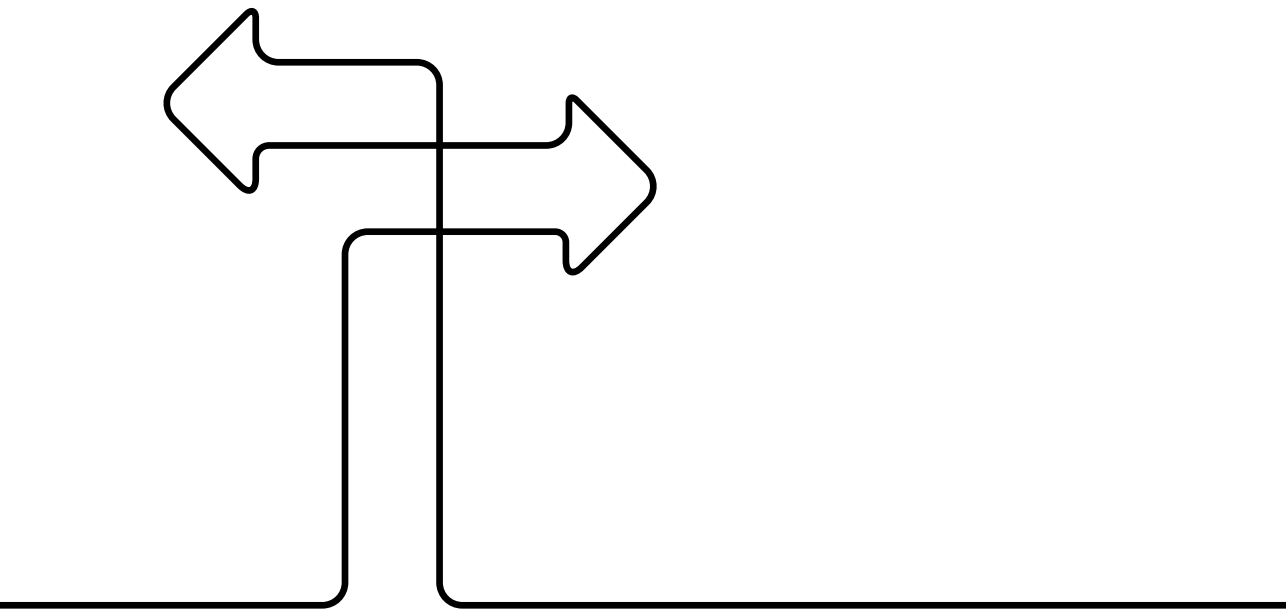
Triage systems and PEWS use data collected during the triage process. These data include pre-defined signs and symptoms that are reported by patients or care-givers. Nurses' clinical impression has been shown to provide additional value above these established predictors of patient urgency. Further research is needed to discover which patient characteristics form the foundation of nurses' clinical impression. In children with

underlying illnesses, parents may already have experienced multiple disease episodes as well, including hospital or even ICU admissions. These parents may be well able to judge the clinical condition of their child and their impression may be a promising predictor for urgency in future studies.[139]

The role of the clinician-researcher

Increasingly advanced possibilities in scientific research result in a growing need for interdisciplinary collaboration. A successful research project requires a relevant research question, full use the possibilities of EHR data, selection of the appropriate study design and application of the most up-to-date statistical methods. This challenge can only be met by multidisciplinary teams consisting of clinicians, informaticians, epidemiologists and statisticians.

However, not all clinicians will be aware of the possibilities of EHR data, nor of the statistical techniques that are available to solve clinical questions. Clinician-researchers are needed to bridge this gap between clinicians and methodologists. They can support the translation of real-world clinical problems into research questions, translate complex methodological concept into clinical research designs, but also translate research results into practice in implementation programmes or guidelines and thus disseminate evidence-based treatments into routine clinical services.[248-250] Through scientifically innovative translational research, clinician-researchers have the ability to translate research findings into health benefits that directly have an impact on patient care.



Chapter 13

Summaries

SUMMARY

In **Chapter 1**, the rationale for this thesis is outlined. Triage systems are used in emergency departments (EDs) as a quick assessment to prioritize patients and ensure they are seen in order of clinical need, rather than in order of attendance. Research on currently used triage systems is limited and often had several methodological limitations. Research is needed to improve available triage systems for children, by reducing undertriage (incorrectly classifying high urgent patients as low urgent), as well as overtriage (incorrectly classifying low urgent patients as high urgent). To this end, existing triage systems, such as the Manchester Triage System (MTS) can be modified, or new tools for the triage assessment can be developed. To improve the first assessment of children presenting at the ED, the TrIAGE project (TRiage Improvements Across General Emergency departments) was established. This is a prospective observational study in five EDs (Austria, Netherlands, Portugal, United Kingdom) based on electronic health record data

The first part of this thesis addresses the performance of currently used triage systems. We propose as minimum requirement for a triage system that it accurately identifies patients in need of admission to the intensive care unit (ICU). Therefore, in **Chapter 2**, we assess the ability of the MTS to correctly identify patients in the ED that require ICU admission. This study was conducted in a large cohort of 50,062 consecutive ED visits in the Erasmus MC- Sophia. Of the 830 children admitted to the ICU during the study period, 238 (28.7%) were undertriaged. Sensitivity of high MTS urgency levels to detect ICU admission was 71% (95% confidence interval (CI) 68%-74%) and specificity 85% (95%CI 85%-85%). Children younger than the age of 3 months, children presenting with medical problems, and children with underlying chronic conditions were at highest risk of undertriage, as well as children referred by emergency medical services or medical specialists, and children presenting during evening and night shifts. These results suggest that the MTS misclassifies a substantial number of children who require ICU admission. Modifications targeted at high risk groups such as young children and children with a comorbid condition could possibly improve safety of the MTS.

To obtain more insight in the validity of the MTS in emergency care, we conducted a prospective observational study of 288,663 patients from three European EDs (**Chapter 3**). In this study, validity of the MTS was determined in the general population of patients attending the ED, specifically for children and elderly, and for commonly used MTS flowcharts and discriminators. Sensitivity of the MTS in the three hospitals ranged from 0.47 (95%CI 0.44-0.49) to 0.87 (95%CI 0.85-0.90), and specificity from 0.84 (95%CI 0.84-0.84) to 0.94 (95%CI 0.94-0.94) for the triage of adult patients. In children, sensitivity ranged from 0.65 (95%CI 0.61-0.70) to 0.83 (95%CI 0.79-0.87), and specificity from 0.83

(95%CI 0.82-0.83) to 0.89 (95%CI 0.88-0.90). There was substantial heterogeneity between the settings, but overall performance was lowest in the young and elderly patients.

Chapter 4 provides an overview of the available evidence on the performance of triage systems in emergency care. Through a systematic review, we identified sixty-six eligible studies evaluating thirty-three different triage systems. Only three triage systems, the Canadian Triage and Acuity Scale, the Emergency Severity Index, and the Manchester Triage System, had multiple evaluations using the same reference standard, and thus could be compared. Overall validity to identify high and low urgency patients of each of these three triage system was moderate to good, but performance was highly variable between studies. In a subgroup analysis, no clear association was found between ED patient volume or case-mix severity of illness and triage systems' performance. These results emphasize that some important research questions still remain: what determinants influence triage systems' performance and how can the performance of existing triage systems be improved.

In **Chapter 5,** performance measures for the assessment of modifications to triage systems are evaluated. Commonly used performance measures do not take into account the specific features of ED triage systems, including their ordinal nature and the different clinical consequences of the different directions of misclassification. We identified available performance measures based on a literature review and expert knowledge. Their properties are illustrated in a case study where we simulated two triage modifications of an ED triage system in a prospective cohort of 14,485 paediatric ED visits. Based on a systematic appraisal of each performance measure's strengths and weaknesses, we propose that decision curves and diagnostic accuracy measures should be used in a dichotomized analysis, and the triage-weighted kappa and Nagelkerke's R^2 in an ordinal approach, when comparing modifications of triage systems.

Part II of this thesis is aimed at improving the MTS. **Chapter 6** describes a study embedded in the TrIAGE project, on the development and validation of vital signs-based modifications to the MTS. In this study, we derived new MTS discriminators based on heart rate, respiratory rate, and capillary refill time for specific subgroups of presentational flowcharts. The optimal cut-off value for each vital sign was determined. Based on the analyses, we propose six new discriminators for children <1 year and ≥1 year: "Very abnormal respiratory rate", "Abnormal heart rate", and "Abnormal respiratory rate", with optimal cut-offs, and specific subgroups of flowcharts. Application of the modified MTS reclassified 744 patients (2.5%). Sensitivity increased from 0.66 (95%CI 0.60-0.72) to 0.71 (95%CI 0.66-0.75) for high urgency patients and from 0.67 (95%CI 0.54-0.76) to 0.70 (95%CI 0.58-0.80) for high and intermediate urgency patients. Specificity decreased from 0.90 (95%CI 0.86-0.93) to 0.89 (95%CI 0.85-0.92) for high and 0.66 (95%CI

0.52-0.78) to 0.63 (95%CI 0.50-0.75) for high and intermediate urgency patients. Overall performance improved (R^2 0.199 versus 0.204). We propose to include these evidence-based modifications in the MTS.

Part III of this thesis, explores additional tools in the first assessment of children. **Chapter 7** assesses the diagnostic accuracy of nurses' clinical impression that a child appears ill, for the recognition of children with serious illness at the ED, in a cohort of 6,390 children. Sensitivity of nurses' clinical impression for the recognition of patients requiring ICU admission was 0.70 (95%CI 0.62-0.76), and specificity 0.81 (95%CI 0.80-0.82). Sensitivity for hospital admission was 0.48 (95%CI 0.45-0.51), and specificity 0.88 (95%CI 0.87-0.88). When adjusted for age, gender, triage urgency and abnormal vital signs, nurses' impression remained significantly associated with ICU (odds ratio (OR) 4.54; 95%CI 3.09 to 6.66), and hospital admission (OR 4.00; 95%CI 3.40 to 4.69). These results demonstrate that the overall clinical impression of experienced nurses at the ED is on its own not an accurate predictor of serious illness in children. It does, however, provide additional information above some well-established and objective predictors of illness severity.

The subsequent two chapters assess the value of hypotension in the recognition of serious illness in children, and describe reference ranges for low blood pressure. In **Chapter 8,** we determined the association between hypotension and serious illness in a large cohort of 10,698 children with measured blood pressure that visited the ED of the Erasmus MC-Sophia. Because paediatric guidelines provide different definitions of low blood pressure, we used three clinical cut-offs to define hypotension: the cut-offs from the Advanced Paediatric Life Support (APLS), the Paediatric Advanced Life Support (PALS)/septic shock screening tool, and the Parshuram's Paediatric Early Warning Score. Hypotension, as a sole predictor, had a significant association with ICU admission (range OR 2.56–5.27) and hospital admission (range OR 1.46–2.66). The association between hypotension and serious illness remained significant after adjustment for tachycardia, and when blood pressure and heart rate were combined in the Shock Index (ratio heart rate : blood pressure). Hypotension showed low sensitivity (range 0.05–0.12) and high specificity (range 0.95–0.99) for ICU and hospital admission, limiting its clinical relevance. These results suggest that routine blood pressure measurement for detection of hypotension is of limited value in all children attending the ED. **Chapter 9** describes a systematic review that aims to identify evidence-based reference values for low blood pressure and compare these with existing definitions for hypotension. Fourteen studies including population-based centiles (first to fifth centile) for non-invasive systolic blood pressure in healthy children < 18 years were included. Existing clinical guidelines for hypotension were identified in international guidelines and textbooks, and show large variability and low to moderate agreement with population-based lower centiles. These results highlight that evidence-based reference values for hypotension for children in the ED are lacking.

Chapter 10 describes the development and validation a novel paediatric early warning score for use in the ED, the ED-PEWS. The ED-PEWS includes patient age and six physiologic parameters: heart rate, respiratory rate, oxygen saturation, consciousness, capillary refill time, and work of breathing. The ED-PEWS showed a cross-validated c-statistic of 0.86 (95% prediction interval 0.82-0.89) for high urgency and 0.67 (0.61-0.73) for high or intermediate urgency patients. A cut-off of score of ≥15 was useful for identifying high urgency patients with a specificity of 0.90 (95%CI 0.87-0.92) while a cut-off score of <6 was useful for identifying low urgency patients with a sensitivity of 0.83 (95%CI 0.81-0.85). The ED-PEWS is the first PEWS derived and validated fully based on statistical methods and therefore aims to represent the optimal score based on currently available evidence.

In **Chapter 11,** the ED-PEWS is validated for children with underlying chronic conditions, in a subset of the TrIAGE cohort. The ED-PEWS had a similar performance in children with comorbidity compared to previously healthy children. Adding the ED-PEWS to the existing triage system, the MTS, for children with comorbidity improved triage, except in the setting with relatively few high urgency patients. Undertriage decreased while overtriage only slightly increased. Overall performance according to Nagelkerke's R^2 improved from 15.3% of the original MTS to 16.0% of the MTS with the ED-PEWS added for children with complex comorbidity, and 16.0% added for children with any comorbidity.

This thesis ends with a general discussion in **Chapter 12**. Different aspects of the TriAGE project are addressed related to how it aims to overcome the challenges of previous triage studies. These include the use of routinely collected electronic health record data, application of appropriate performance measures, the definition of a three level reference standard for patient urgency, and the combination of clinical and methodological strategies in the analysis. Strengths and limitations of the project are discussed. In the last part of the discussion, we comment on future perspectives for triage studies. Future research projects should address the development of a large, potentially nationwide database including data from ED visits. Clinical strategies and statistical modelling should be combined to provide evidence-based recommendations. Innovations in the first assessment of children are needed, particularly the implementation of PEWS and other decision rules in EHRs, the possibility to provide automatic generation of practice recommendations to triage nurses, and the exploration of new predictors of patient urgency. Finally, clinician-researchers should bridge the gap between clinicians and methodologists, as they can translate real-world clinical problems into research questions, and translate research results into practice in implementation programmes or guidelines.

NEDERLANDSE SAMENVATTING

In **Hoofdstuk 1** wordt de rationale van dit proefschrift beschreven. Triage systemen worden gebruikt op de spoedeisende hulp (SEH) als een snelle beoordeling om patiënten te prioriteren en ervoor te zorgen dat ze worden geholpen op basis van urgentie en niet op volgorde van binnenkomst. Er is slechts beperkt onderzoek gedaan naar de momenteel veelgebruikte triagesystemen en bestaande studies hebben vaak meerdere methodologische beperkingen. Onderzoek is noodzakelijk om beschikbare triagesystemen voor kinderen te verbeteren, door zowel ondertriage (het incorrect classificeren van hoog urgente patiënten als laag urgent) als overtriage (het incorrect classificeren van laag urgente patiënten als hoog urgent) te verminderen. Dit kan door bestaande triage systemen, zoals het Manchester Triage Systeem (MTS) aan te passen, of door nieuwe instrumenten voor triage te ontwikkelen. Het TrIAGE project (TRiage Improvements Across General Emergency departments) is opgericht om de eerste beoordeling van kinderen op de SEH te verbeteren. Het is een prospectieve observationele studie op vijf SEHs (Nederland, Oostenrijk, Portugal, Verenigd Koninkrijk), gebaseerd op gegevens uit het elektronisch patiëntendossier.

Het eerste deel van dit proefschrift betreft de validiteit van in de praktijk gebruikte triagesystemen. Wij stellen dat een minimale vereiste voor ieder triagesysteem is, dat het patiënten die Intensive Care (IC) behoeftig zijn accuraat identificeert. Daarom beoordelen we in **Hoofdstuk 2** of het MTS op de SEH correct patiënten kan identificeren die IC opname nodig hebben. Deze studie werd uitgevoerd in een groot cohort van 50,062 opeenvolgende SEH bezoeken in het Erasmus MC- Sophia. Van de 830 kinderen opgenomen op de IC tijdens de onderzoeksperiode waren 238 (28.7%) ondergetrieerd door het MTS. Sensitiviteit van hoge MTS urgentie om IC opname te detecteren was 71% (95% betrouwbaarheidsinterval (BI) 68%-74%) en specificiteit 85% (95%BI 85%-85%). Kinderen jonger dan 3 maanden, kinderen die zich presenteren met medische problemen en kinderen met onderliggende chronische aandoeningen hadden het hoogste risico op ondertriage, samen met kinderen verwezen door de alarmcentrale of een medisch specialist en kinderen die zich presenteren tijdens de avond- en nachtdiensten. Deze resultaten suggereren dat het MTS een substantieel aantal kinderen die IC-behoeftig zijn niet correct classificeert. Aanpassingen die zich richten op hoog-risicogroepen zoals jonge kinderen en kinderen met comorbiditeit kan mogelijk de veiligheid van het MTS verbeteren.

Om meer inzicht te krijgen in de validiteit van het MTS in de spoedzorg, verrichtten we een prospectieve observationele studie van 288,663 patiënten van drie Europese SEHs (**Hoofdstuk3**). In deze studie bepaalden we de validiteit van het MTS in de algemene populatie van patiënten die de SEH bezoeken, met specifieke aandacht voor kinderen

en ouderen en veel gebruikte MTS stroomschema's en discriminatoren. Sensitiviteit van het MTS in de drie ziekenhuizen varieerde van 0.47 (95%BI 0.44-0.49) tot 0.87 (95%BI 0.85-0.90) en specificiteit van 0.84 (95%BI 0.84-0.84) tot 0.94 (95%BI 0.94-0.94) voor de triage van volwassenen. Voor kinderen varieerde de sensitiviteit van 0.65 (95%BI 0.61-0.70) tot 0.83 (95%BI 0.79-0.87) en specificiteit van 0.83 (95%BI 0.82-0.83) tot 0.89 (95%BI 0.88-0.90). Er was aanzienlijke heterogeniteit tussen de verschillende SEHs, maar in het algemeen was de validiteit het laagst voor de jonge en oudere patiënten.

Hoofdstuk 4 geeft een overzicht van het beschikbare bewijs met betrekking tot de validiteit van triagesystemen op de SEH. Middels een systematisch literatuuronderzoek, vonden we zesenzestig geschikte studies die drieëndertig verschillende triagesystemen evalueerden. Slechts drie triagesystemen, de Canadian Triage and Acuity Scale, de Emergency Severity Index en het Manchester Triage Systeem werden in meerdere studies geëvalueerd middels dezelfde referentiestandaard en konden dus vergeleken worden. De validiteit om hoog en laag urgente patiënten te identificeren was van elk van deze drie triagesystemen goed, maar er waren grote verschillen tussen de verschillende studies. In een subgroep analyse werd geen duidelijk verband gevonden tussen SEH volume of gemiddelde ernst van de SEH bezoeken en de validiteit van het triagesysteem. Deze resultaten benadrukken dat er nog belangrijke onderzoeksvragen onbeantwoord blijven: welke determinanten beïnvloeden de werking van triagesystemen en hoe kan de validiteit van triagesystemen verbeterd worden.

In **Hoofdstuk 5** worden maten voor het beoordelen van aanpassingen aan triagesystemen geëvalueerd. Veelgebruikte maten houden meestal geen rekening met de specifieke eigenschappen van triagesystemen, zoals de ordinale structuur en de verschillende klinische consequenties van de verschillende soorten misclassificatie. We identificeerden gebruikte maten door middel van een literatuuronderzoek en de kennis van experts. Eigenschappen illustreerden we middels een casus waarbij we twee aanpassingen aan een triagesysteem simuleerden in een prospectief cohort van 14,485 SEH bezoeken van kinderen. Gebaseerd op een systematische beoordeling van de sterke en zwakke punten van elke maat, doen wij het voorstel om bij het vergelijken van aanpassingen aan een triagesysteem besliscurves en maten van diagnostische accuraatheid te gebruiken in een dichotome analyse, en de triage-gewogen kappa en Nagelkerke's R^2 bij een ordinale benadering.

Deel II van dit proefschrift is gericht op het verbeteren van het MTS. **Hoofdstuk 6** beschrijft een studie, als onderdeel van het TrIAGE project, waarin aanpassingen van het MTS gebaseerd op vitale parameters ontwikkeld en gevalideerd worden. In deze studie ontwikkelden we nieuwe MTS discriminatoren op basis van hartfrequentie, ademhalingsfrequentie en capillaire refill tijd voor specifieke subgroepen van MTS

stroomschema's. Voor elk van de vitale parameters werd de optimale afkapwaarde bepaald. Op basis van de analyses, stellen we zes nieuwe discriminatoren voor, voor kinderen <1 jaar en ≥1 jaar: "Zeer hoge ademhalingsfrequentie", "Hoge hartfrequentie", en "Hoge ademhalingsfrequentie", met optimale afkapwaardes en specifieke subgroepen van MTS stroomschema's. Toepassing van het gewijzigde MTS zorgde voor een andere triage classificatie van 744 patiënten (2.5%). De sensitiviteit nam toe van 0.66 (95%BI 0.60-0.72) tot 0.71 (95%BI 0.66-0.75) voor hoog urgente patiënten en van 0.67 (95%BI 0.54-0.76) tot 0.70 (95%BI 0.58-0.80) voor hoog en medium urgente patiënten. Specificiteit nam af van 0.90 (95%BI 0.86-0.93) tot 0.89 (95%BI 0.85-0.92) voor hoog en 0.66 (95%BI 0.52-0.78) tot 0.63 (95%BI 0.50-0.75) voor hoog en medium urgente patiënten. De algemene validiteit verbeterde (R2 0.199 versus 0.204). We stellen voor om deze evidence-based modificaties toe te passen in het MTS.

Deel III van dit proefschrift gaat over andere hulpmiddelen voor de eerste beoordeling van kinderen. **Hoofdstuk 7** onderzoekt de diagnostische accuraatheid van de klinische indruk van verpleegkundigen dat een kind ziek oogt, voor de herkenning van acuut zieke kinderen op de SEH, in een cohort van 6,390 kinderen. Sensitiviteit van de klinische blik van verpleegkundigen voor het herkennen van patiënten met IC-behoefte was 0.70 (95%BI 0.62-0.76), en specificiteit 0.81 (95%BI 0.80-0.82). Sensitiviteit voor ziekenhuisopname was 0.48 (95%BI 0.45-0.51) en specificiteit 0.88 (95%BI 0.87-0.88). Gecorrigeerd voor leeftijd, geslacht, triage urgentie en afwijkende vitale parameters, bleef de klinische blik van verpleegkundigen significant geassocieerd met IC (odds ratio (OR) 4.54; 95%BI 3.09 tot 6.66) en ziekenhuisopname (OR 4.00; 95%BI 3.40 tot 4.69). Deze resultaten laten zien dat de klinische blik van ervaren verpleegkundigen op de SEH op zichzelf geen accurate voorspeller is voor ernstige ziekte bij kinderen. Het heeft echter wel toegevoegde waarde boven een aantal bekende en objectieve voorspellers van ziekte ernst.

De volgende twee hoofdstukken onderzoeken de waarde van hypotensie bij de herkenning van acute zieke kinderen en beschrijven de referentiewaarden van lage bloeddruk. In **Hoofdstuk 8** onderzoeken we de associatie tussen hypotensie en ernstige ziekte in een groot cohort van 10,698 kinderen met gemeten bloeddruk die de SEH van het Erasmus MC-Sophia bezochten. Omdat kindergeneeskundige richtlijnen verschillende definities van lage bloeddruk aanhouden, gebruikten we drie klinische afkapwaarden om hypotensie vast te stellen: de afkapwaarden van de Advanced Paediatric Life Support (APLS), de Paediatric Advanced Life Support (PALS)/septic shock screening tool, en Parshuram's Paediatric Early Warning Score. Hypotensie als enige voorspeller had een significante associatie met IC opname (range OR 2.56 tot 5.26) en ziekenhuisopname (range OR 1.46 tot 2.66). De associatie tussen hypotensie en ernstige ziekte bleef significant na correctie voor tachycardie en als bloeddruk en hartfrequentie werden gecombineerd in de Shock Index (ratio van hartfrequentie : bloeddruk). Hypotensie had een lage sensitiviteit (range

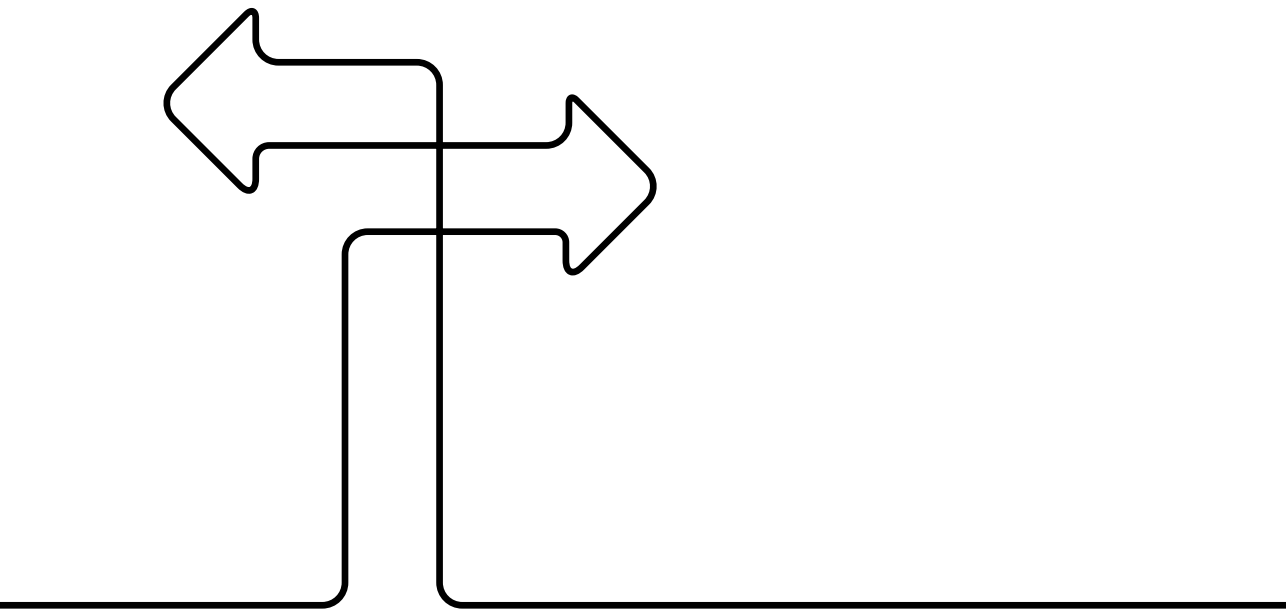
0.05-0.12) en hoge specificiteit (range 0.95-0.99) voor IC en ziekenhuisopname, waardoor de klinische relevantie ervan beperkt is. Deze resultaten suggereren dat routinematige bloeddrukmeting voor het vaststellen van hypotensie op de SEH van beperkte waarde is bij kinderen. **Hoofdstuk 9** beschrijft een systematisch literatuuronderzoek met als doel om evidence-based referentiewaarden te identificeren voor lage bloeddruk en om deze te vergelijken met bestaande definities van hypotensie. Veertien studies met percentielen (eerste tot vijfde) van niet-invasieve systolische bloeddruk bij gezonde kinderen <18 jaar werden geïncludeerd. Bestaande klinische richtlijnen voor hypotensie werden geïdentificeerd uit internationale richtlijnen en tekstboeken en tonen grote variabiliteit en lage tot matige overeenkomst met de op populatie gebaseerde onderste percentielen. Deze resultaten benadrukken dat evidence-based referentiewaarden voor hypotensie bij kinderen ontbreken.

Hoofdstuk 10 beschrijft de ontwikkeling en validatie van een nieuw vroegtijdig waarschuwingssysteem voor gebruik bij kinderen op de SEH, de ED-PEWS (Emergency Department – Paediatric Early Warning Score). De ED-PEWS bestaat uit leeftijd en zes fysiologische parameters: hartfrequentie, ademhalingsfrequentie, zuurstofsaturatie, bewustzijn, capillaire refill tijd en ademhalingsarbeid. De ED-PEWS liet een kruis-gevalideerde c statistiek zien van 0.86 (95% predictie interval 0.82-0.89) voor hoog en 0.67 (0.61-0.73) voor hoog tot medium urgente patiënten. Een cutoff score ≥15 was bruikbaar om hoog urgente patiënten te identificeren met een specificiteit van 0.90 (95%BI 0.87-0.92) terwijl een cutoff score <6 bruikbaar was voor het identificeren van laag urgente patiënten met een sensitiviteit van 0.83 (95%BI 0.81-0.85). De ED-PEWS is de eerste PEWS die volledig gederiveerd en gevalideerd is middels statistische methoden en daarmee poogt om de optimale score te zijn gebaseerd op huidig beschikbaar bewijs.

In **Hoofdstuk 11** wordt de ED-PEWS gevalideerd voor kinderen met onderliggende chronische aandoeningen in een subgroep van het TrIAGE cohort. De validiteit van de ED-PEWS was vergelijkbaar in kinderen met comorbiditeit en voorheen gezonde kinderen. Het toevoegen van de ED-PEWS aan het bestaande triagesysteem, het MTS, bij kinderen met comorbiditeit verbeterde de triage, behalve in de setting met relatief weinig hoog urgente patienten. De ondertriage nam af terwijl overtriage nauwelijks toenam. Validiteit volgens Nagelkerke's R^2 verbeterde van 15.3% voor het originele MTS naar 16.0% voor het MTS waarbij de ED-PEWS werd toegevoegd voor kinderen met complexe comorbiditeit en 16.0% voor kinderen met zowel complexe als niet-complexe comorbiditeit.

Dit proefschrift eindigt met een algemene discussie in **Hoofdstuk 12**. Hierin worden verschillende aspecten van het TrIAGE project besproken die pogen om de beperkingen van eerder triage onderzoek aan te pakken. Deze omvatten onder ander het gebruik van routinematig verzamelde gegevens uit het elektronisch patiëntendossier, het gebruik

van passende maten om validiteit te beschrijven, de definitie van een referentiestandaard bestaande uit drie categorieën van urgentie en de combinatie van klinische en methodologische strategieën in de analyse. De sterke punten en beperkingen van het project worden besproken. In het laatste deel van de discussie bespreken we verschillende toekomstperspectieven met betrekking tot triage onderzoek. Toekomstige onderzoeksprojecten zouden zich moeten richten op de ontwikkeling van grote, mogelijk nationale databases met gegevens van SEH bezoeken. Klinische strategieën en statistisch modelleren moeten gecombineerd worden om evidence-based aanbevelingen te kunnen doen. Innovaties in de eerste beoordeling van kinderen zijn noodzakelijk, in het bijzonder de implementatie van PEWS en andere beslisregels in het elektronisch patiëntendossier, de mogelijkheid om geautomatiseerde aanbevelingen te genereren voor triage verpleegkundigen en onderzoek naar nieuwe voorspellers van urgentie. Uiteindelijk zouden clinicus-onderzoekers de brug moeten slaan tussen clinici en methodologen, omdat ze bestaande problemen kunnen vertalen naar onderzoeksvragen en andersom onderzoeksresultaten kunnen vertalen naar de praktijk in implementatie programma's of richtlijnen.



APPENDICES

I. REFERENCES

1. Gebruik acute zorg. https://www.volksgezondheidenzorg.info/onderwerp/acute-zorg/cijfers-context/gebruik-acute-zorg#node-gebruik-seh-naar-leeftijd. Published 2016. Accessed May 21, 2020.
2. STATISTICAL BRIEF #242. https://www.hcup-us.ahrq.gov/reports/statbriefs/sb242-Pediatric-ED-Visits-2015.pdf. Published 2015. Accessed May 21, 2020.
3. Hostetler MA, Mace S, Brown K, Finkler J, Hernandez D, Krug SE, et al. Emergency department overcrowding and children. *Pediatr Emerg Care*. 2007;23(7):507-515.
4. Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. *Ann Emerg Med*. 2008;52(2):126-136.
5. Athey J, Dean JM, Ball J, Wiebe R, Melese-d'Hospital I. Ability of hospitals to care for pediatric emergency patients. *Pediatr Emerg Care*. 2001;17(3):170-174.
6. National Research Council. *Emergency Medical Services for Children*. Washington, DC: The National Academies Press; 1993.
7. Pearson GA, Ward-Platt M, Harnden A, Kelly D. Why children die: avoidable factors associated with child deaths. *Arch Dis Child*. 2011;96(10):927-931.
8. Wolfe I, Thompson M, Gill P, Tamburlini G, Blair M, van den Bruel A, et al. Health services for children in western Europe. *Lancet (London, England)*. 2013;381(9873):1224-1234.
9. FitzGerald G, Jelinek GA, Scott D, Gerdtz MF. Emergency department triage revisited. *Emerg Med J*. 2010;27(2):86-92.
10. Warren DW, Jarvis A, LeBlanc L, Gravel J, Group CNW, Canadian Association of Emergency P, et al. Revisions to the Canadian Triage and Acuity Scale paediatric guidelines (PaedCTAS). *Cjem*. 2008;10(3):224-243.
11. Gilboy NT, P.; Travers, D.A.; Rosenau, A.M.; Eitel, D.R. *Emergency Severity Index, Version 4: Implementation Handbook*. Rockville, MD: Agency for Healthcare Research and Quality; 2005.
12. Mackway-Jones K, Marsden J, Windle J. *Emergency triage*. London: Blackwell Publishing Ltd.; 2006.
13. Zachariasse JM, van der Hagen V, Seiger N, Mackway-Jones K, van Veen M, Moll HA. Performance of triage systems in emergency care: a systematic review and meta-analysis. *BMJ Open*. 2019;9(5):e026471.
14. Mackway-Jones K, Marsden J, Windle J. *Emergency Triage*: John Wiley & Sons, Ltd.; 2014.
15. van Veen M, Teunen-van der Walle VF, Steyerberg EW, van Meurs AH, Ruige M, Strout TD, et al. Repeatability of the Manchester Triage System for children. *Emerg Med J*. 2010;27(7):512-516.
16. Mirhaghi A, Mazlom R, Heydari A, Ebrahimi M. The reliability of the Manchester Triage System (MTS): a meta-analysis. *J Evid Based Med*. 2017;10(2):129-135.
17. Moll HA. Challenges in the validation of triage systems at emergency departments. *J Clin Epidemiol*. 2010;63(4):384-388.
18. Schein RM, Hazday N, Pena M, Ruben BH, Sprung CL. Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest*. 1990;98(6):1388-1392.
19. Buist MD, Jarmolowski E, Burton PR, Bernard SA, Waxman BP, Anderson J. Recognising clinical instability in hospital patients before cardiac arrest or unplanned admission to intensive care. A pilot study in a tertiary-care hospital. *Med J Aust*. 1999;171(1):22-25.
20. Tume L. The deterioration of children in ward areas in a specialist children's hospital. *Nurs Crit Care*. 2007;12(1):12-19.
21. Nijman RG, Vergouwe Y, Thompson M, van Veen M, van Meurs AH, van der Lei J, et al. Clinical prediction model to aid emergency doctors managing febrile children at risk of serious bacterial infections: diagnostic study. *BMJ (Clinical research ed)*. 2013;346:f1706.

22. Schlapbach LJ, Straney L, Bellomo R, MacLaren G, Pilcher D. Prognostic accuracy of age-adapted SOFA, SIRS, PELOD-2, and qSOFA for in-hospital mortality among children with suspected infection admitted to the intensive care unit. *Intensive Care Med*. 2018;44(2):179-188.
23. Seiger N, van Veen M, Steyerberg EW, Ruige M, van Meurs AH, Moll HA. Undertriage in the Manchester triage system: an assessment of severity and options for improvement. *Arch Dis Child*. 2011;96(7):653-657.
24. Lamb B, Sevdalis N. How do nurses make decisions? *Int J Nurs Stud*. 2011;48(3):281-284.
25. Thompson C, Aitken L, Doran D, Dowding D. An agenda for clinical decision making and judgement in nursing research and education. *Int J Nurs Stud*. 2013;50(12):1720-1726.
26. Gilhotra Y, Willis F. Blood pressure measurements on children in the emergency department. *Emerg Med Australas*. 2006;18(2):148-154.
27. Thompson T, Stanford K, Dick R, Graham J. Triage assessment in pediatric emergency departments: a national survey. *Pediatr Emerg Care*. 2010;26(8):544-548.
28. Haque IU, Zaritsky AL. Analysis of the evidence for the lower limit of systolic and mean arterial pressure in children. *Pediatr Crit Care Med*. 2007;8(2):138-144.
29. Seiger N, Moll HA. Triage systems: outcome measures to validate. *Ann Emerg Med*. 2013;61(3):372-373.
30. Roukema J, Steyerberg EW, van Meurs A, Ruige M, van der Lei J, Moll HA. Validity of the Manchester Triage System in paediatric emergency care. *Emerg Med J*. 2006;23(12):906-910.
31. van Veen M, Steyerberg EW, Ruige M, van Meurs AH, Roukema J, van der Lei J, et al. Manchester triage system in paediatric emergency care: prospective observational study. *BMJ (Clinical research ed)*. 2008;337:a1501.
32. Cooke MW, Jinks S. Does the Manchester triage system detect the critically ill? *J Accid Emerg Med*. 1999;16(3):179-181.
33. Chalfin DB, Trzeciak S, Likourezos A, Baumann BM, Dellinger RP, group D-Es. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Crit Care Med*. 2007;35(6):1477-1483.
34. Seiger N, van Veen M, Steyerberg EW, van der Lei J, Moll HA. Accuracy of triage for children with chronic illness and infectious symptoms. *Pediatrics*. 2013;132(6):e1602-1608.
35. van Veen M, Steyerberg EW, Van't Klooster M, Ruige M, van Meurs AH, van der Lei J, et al. The Manchester triage system: improvements for paediatric emergency care. *Emerg Med J*. 2012;29(8):654-659.
36. Nederlandse Vereniging voor Kindergeneeskunde. Zorg voor kwaliteit van zorg voor zieke kinderen - Kwaliteitscriteria voor intensive care voor kinderen in Nederland. Utrecht2012.
37. Manchester Triage Group. *Triage voor de spoedeisende hulp*. 2 ed: Elsevier gezondheidszorg; 2007.
38. Visser IH, Hazelzet JA, Albers MJ, Verlaat CW, Hogenbirk K, van Woensel JB, et al. Mortality prediction models for pediatric intensive care: comparison of overall and subgroup specific performance. *Intensive Care Med*. 2013;39(5):942-950.
39. Slater A, Shann F, Pearson G, Paediatric Index of Mortality Study G. PIM2: a revised version of the Paediatric Index of Mortality. *Intensive Care Med*. 2003;29(2):278-285.
40. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. *Crit Care Med*. 1996;24(5):743-752.
41. Simon TD, Cawthon ML, Stanford S, Popalisky J, Lyons D, Woodcox P, et al. Pediatric medical complexity algorithm: a new method to stratify children by medical complexity. *Pediatrics*. 2014;133(6):e1647-1654.
42. Graff I, Goldschmidt B, Glien P, Bogdanow M, Fimmers R, Hoeft A, et al. The German Version of the Manchester Triage System and its quality criteria--first assessment of validity and reliability. *PLoS One*. 2014;9(2):e88995.
43. Santos AP, Freitas P, Martins HM. Manchester Triage System version II and resource utilisation in the emergency department. *Emergency medicine journal : EMJ*. 2014;31(2):148-152.

44. 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):506-509.
45. Storm-Versloot MN, Ubbink DT, Chin a Choi V, Luitse JS. Observer agreement of the Manchester Triage System and the Emergency Severity Index: a simulation study. *Emerg Med J*. 2009;26(8):556-560.
46. Dosa NP, Boeing NM, Ms N, Kanter RK. Excess risk of severe acute illness in children with chronic health conditions. *Pediatrics*. 2001;107(3):499-504.
47. Cameron P, Jelinek G, Everitt I. *Textbook of paediatric emergency medicine*: Elsevier Health Sciences; 2006.
48. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med*. 2001;345(19):1368-1377.
49. Thompson MJ, Ninis N, Perera R, Mayon-White R, Phillips C, Bailey L, et al. Clinical recognition of meningococcal disease in children and adolescents. *Lancet (London, England)*. 2006;367(9508):397-403.
50. American Academy of Pediatrics Committee on Pediatric Emergency Medicine. Overcrowding crisis in our nation's emergency departments: is our safety net unraveling? *Pediatrics*. 2004;114(3):878-888.
51. Institute Of M. IOM report: the future of emergency care in the United States health system. *Acad Emerg Med*. 2006;13(10):1081-1085.
52. Australasian College for Emergency Medicine. Access Block and Overcrowding in Emergency departments. 2004.
53. NHS England. Transforming urgent and emergency care services in England - Urgent and Emergency Care Review. 2013.
54. Pinto D, Jr., Salgado Pde O, Chianca TC. Predictive validity of the Manchester Triage System: evaluation of outcomes of patients admitted to an emergency department. *Rev Lat Am Enfermagem*. 2012;20(6):1041-1047.
55. 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):822-829.
56. Pinto D, Lunet N, Azevedo A. Sensitivity and specificity of the Manchester Triage System for patients with acute coronary syndrome. *Rev Port Cardiol*. 2010;29(6):961-987.
57. Providencia R, Gomes PL, Barra S, Silva J, Seca L, Antunes A, et al. Importance of Manchester Triage in acute myocardial infarction: impact on prognosis. *Emerg Med J*. 2011;28(3):212-216.
58. Manchester Triage Groep. *Triage voor de spoedeisende hulp*. 2 ed: Elsevier gezondheidszorg; 2007.
59. Grupo Português de Triagem. *Triagem no serviço de urgência*. Amadora: Blackwell Publishing Ltd; 2009.
60. Seiger N. Improvements of Paediatric Triage at the Emergency Department: Erasmus University Rotterdam; 2014.
61. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. *QJM*. 2001;94(10):521-526.
62. Platts-Mills TF, Travers D, Biese K, McCall B, Kizer S, LaMantia M, et al. Accuracy of the Emergency Severity Index triage instrument for identifying elder emergency department patients receiving an immediate life-saving intervention. *Acad Emerg Med*. 2010;17(3):238-243.
63. Phillips S, Rond PC, 3rd, Kelly SM, Swartz PD. The failure of triage criteria to identify geriatric patients with trauma: results from the Florida Trauma Triage Study. *J Trauma*. 1996;40(2):278-283.
64. Staudenmayer KL, Hsia RY, Mann NC, Spain DA, Newgard CD. Triage of elderly trauma patients: a population-based perspective. *J Am Coll Surg*. 2013;217(4):569-576.
65. Nakamura Y, Daya M, Bulger EM, Schreiber M, Mackersie R, Hsia RY, et al. Evaluating age in the field triage of injured persons. *Ann Emerg Med*. 2012;60(3):335-345.

66. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*. 2004;140(3):189-202.
67. Grouse AI, Bishop RO, Bannon AM. The Manchester Triage System provides good reliability in an Australian emergency department. *Emerg Med J*. 2009;26(7):484-486.
68. 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):431-434.
69. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health technology assessment (Winchester, England)*. 2007;11(50):iii, ix-51.
70. Blunt I, Edwards N, Merry L. What's behind the A&E 'crisis'? *Briefing*: Nuffield Trust; 2015.
71. Australasian College for Emergency Medicine. Access Block and Overcrowding in Emergency departments. . 2004.
72. Institute of Medicine. Hospital-based emergency care: at the breaking point. . 2013.
73. Christ M, Grossmann F, Winter D, Bingisser R, Platz E. Modern triage in the emergency department. *Dtsch Arztebl*. 2010;107(50):892-898.
74. Farrohknia N, Castren M, Ehrenberg A, Lind L, Oredsson S, Jonsson H, et al. Emergency department triage scales and their components: a systematic review of the scientific evidence. *Scand J Trauma Resusc Emerg Med*. 2011;19:42.
75. Parenti N, Ferrara L, Bacchi Reggiani ML, Sangiorgi D, Lenzi T. Reliability and validity of two four-level emergency triage systems. *Eur J Emerg Med*. 2009;16(3):115-120.
76. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *Jama*. 2000;283(15):2008-2012.
77. World Bank Group. World Bank Country and Lending Groups. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups. Accessed October 5th, 2017.
78. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
79. Kuriyama A, Urushidani S, Nakayama T. Five-level emergency triage systems: variation in assessment of validity. *Emerg Med J*. 2017.
80. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008.
81. IBM Corp. IBM SPSS Statistics for Windows. 21 ed. Armonk, NY 2011.
82. Review Manager (RevMan). 5.3 ed. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.
83. Chang YC, Ng CJ, Wu CT, Chen LC, Chen JC, Hsu KH. Effectiveness of a five-level Paediatric Triage System: An analysis of resource utilisation in the emergency department in Taiwan. *Emerg Med J*. 2012.
84. Chi CH, Huang CM. Comparison of the Emergency Severity Index (ESI) and the Taiwan triage system in predicting resource utilization. *J Formos Med Assoc*. 2006;105(8):617-625.
85. Maleki M, Fallah R, Riahi L, Delavari S, Rezaei S. Effectiveness of Five-Level Emergency Severity Index Triage System Compared With Three-Level Spot Check: An Iranian Experience. *Arch Trauma Res*. 2015;4(4):e29214.
86. Ng CJ, Hsu KH, Kuan JT, Chiu TF, Chen WK, Lin HJ, et al. Comparison between Canadian Triage and Acuity Scale and Taiwan Triage System in emergency departments. *J Formos Med Assoc*. 2010;109(11):828-837.
87. Ng CJ, Yen ZS, Tsai JCH, Chen LC, Lin SJ, Sang YY, et al. Validation of the Taiwan triage and acuity scale: A new computerised five-level triage system. *Emerg Med J*. 2011;28(12):1026-1031.

88. Yuksen C, Sawatmongkornkul S, Suttabuth S, Sawanyawisuth K, Sittichanbuncha Y. Emergency severity index compared with 4-level triage at the emergency department of Ramathibodi University Hospital. *Asian Biomed*. 2016;10(2):155-161.

89. Travers DA, Waller AE, Bowling JM, Flowers D, Tintinalli J. Five-level triage system more effective than three-level in tertiary emergency department. *J Emerg Nurs*. 2002;28(5):395-400.

90. Gouin S, Gravel J, Amre DK, Bergeron S. Evaluation of the Paediatric Canadian Triage and Acuity Scale in a pediatric ED. *Am J Emerg Med*. 2005;23(3):243-247.

91. Storm-Versloot MN, Ubbink DT, Kappelhof J, Luitse JSK. 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):822-829.

92. van der Wulp I, Schrijvers AJP, van Stel HF. Predicting admission and mortality with the Emergency Severity Index and the Manchester Triage System: a retrospective observational study. *Emergency Medicine Journal*. 2009;26(7):506-509.

93. Worster A, Fernandes CM, Eva K, Upadhye S. Predictive validity comparison of two five-level triage acuity scales. *Eur J Emerg Med*. 2007;14(4):188-192.

94. Ghafarypour-Jahrom M, Taghizadeh M, Heidari K, Derakhshanfar H. Validity and Reliability of the Emergency Severity Index and Australasian Triage System in Pediatric Emergency Care of Mofid Children's Hospital in Iran. *Bull Emerg Trauma*. 2018;6(4):329-333.

95. Aeimchanbanjong K, Pandee U. Validation of different pediatric triage systems in the emergency department. *World J Emerg Med*. 2017;8(3):223-227.

96. Azeredo TR, Guedes HM, Rebelo de Almeida RA, Chianca TC, Martins JC. Efficacy of the Manchester Triage System: a systematic review. *Int Emerg Nurs*. 2015;23(2):47-52.

97. Souza CC, Araujo FA, Chianca TC. [Scientific Literature on the Reliability and Validity of the Manchester Triage System (MTS) Protocol: A Integrative Literature Review]
Producao cientifica sobre a validade e confiabilidade do Protocolo de Manchester: revisao integrativa da literature. *Rev Esc Enferm USP*. 2015;49(1):144-151.

98. van Veen M, Moll HA. Reliability and validity of triage systems in paediatric emergency care. *Scand J Trauma Resusc Emerg Med*. 2009;17:38.

99. de Magalhaes-Barbosa MC, Robaina JR, Prata-Barbosa A, Lopes CS. Validity of triage systems for paediatric emergency care: a systematic review. *Emerg Med J*. 2017.

100. Hansoti B, Jenson A, Keefe D, De Ramirez SS, Anest T, Twomey M, et al. Reliability and validity of pediatric triage tools evaluated in Low resource settings: a systematic review. *BMC Pediatr*. 2017;17(1):37.

101. Jenson A, Hansoti B, Rothman R, de Ramirez SS, Lobner K, Wallis L. Reliability and validity of emergency department triage tools in low- and middle-income countries: a systematic review. *Eur J Emerg Med*. 2018;25(3):154-160.

102. Turner N, Leroy P. *Advanced Paediatric Life Support*. Amsterdam: Reed Business; 2011.

103. van de Maat J, Jonkman H, van de Voort E, Mintegi S, Gervaix A, Bressan S, et al. Measuring vital signs in children with fever at the emergency department: an observational study on adherence to the NICE recommendations in Europe. *Eur J Pediatr*. 2020.

104. Pacella CB, Yealy DM. Are we asking the right triage questions? *Ann Emerg Med*. 2013;61(1):33-34.

105. Cooper RJ. Emergency department triage: why we need a research agenda. *Ann Emerg Med*. 2004;44(5):524-526.

106. 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):128-138.

107. Biesheuvel CJ, Vergouwe Y, Steyerberg EW, Grobbee DE, Moons KG. Polytomous logistic regression analysis could be applied more often in diagnostic research. *J Clin Epidemiol*. 2008;61(2):125-134.

108. Van Calster B, Van Belle V, Vergouwe Y, Steyerberg EW. Discrimination ability of prediction models for ordinal outcomes: relationships between existing measures and a new measure. *Biometrical journal Biometrische Zeitschrift*. 2012;54(5):674-685.
109. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574.
110. van der Wulp I, van Stel HF. Adjusting weighted kappa for severity of mistriage decreases reported reliability of emergency department triage systems: a comparative study. *J Clin Epidemiol*. 2009;62(11):1196-1201.
111. Merkle EC, Steyvers M. Choosing a Strictly Proper Scoring Rule. *Decis Anal*. 2013;10(4):292-304.
112. Gneiting T, Raftery AR. Strictly Proper Scoring Rules, Prediction, and Estimation. *J Am Stat Assoc*. 2007;102(477):359-378.
113. Seiger N, van Veen M, Almeida H, Steyerberg EW, van Meurs AH, Carneiro R, et al. Improving the Manchester Triage System for pediatric emergency care: an international multicenter study. *PLoS One*. 2014;9(1):e83267.
114. 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):678-689.
115. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
116. Baumann MR, Strout TD. Evaluation of the Emergency Severity Index (version 3) triage algorithm in pediatric patients. *Acad Emerg Med*. 2005;12(3):219-224.
117. Gravel J, Manzano S, Arsenault M. Validity of the Canadian Paediatric Triage and Acuity Scale in a tertiary care hospital. *Cjem*. 2009;11(1):23-28.
118. Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The Emergency Severity Index triage algorithm version 2 is reliable and valid. *Acad Emerg Med*. 2003;10(10):1070-1080.
119. 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.
120. Wolfsdorf JI, Glaser N, Agus M, Fritsch M, Hanas R, Rewers A, et al. ISPAD Clinical Practice Consensus Guidelines 2018: Diabetic ketoacidosis and the hyperglycemic hyperosmolar state. *Pediatr Diabetes*. 2018;19 Suppl 27:155-177.
121. Lambert V, Matthews A, MacDonell R, Fitzsimons J. Paediatric early warning systems for detecting and responding to clinical deterioration in children: a systematic review. *BMJ Open*. 2017;7(3):e014497.
122. Irwin AD, Grant A, Williams R, Kolamunnage-Dona R, Drew RJ, Paulus S, et al. Predicting Risk of Serious Bacterial Infections in Febrile Children in the Emergency Department. *Pediatrics*. 2017;140(2).
123. National Institute for Health and Care Excellence (NICE). Meningitis (bacterial) and meningococcal septicaemia in under 16s: recognition, diagnosis and management. 2015.
124. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *J Clin Epidemiol*. 2015;68(2):134-143.
125. Zachariasse JM, Nieboer D, Maconochie IK, Smit FJ, Alves CF, Greber-Platzer S, et al. Development and validation of a Paediatric Early Warning Score for use in the emergency department: a multicentre study. *Lancet Child Adolesc Health*. 2020;[Accepted for publication].
126. Sterne JA, 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 (Clinical research ed)*. 2009;338:b2393.
127. Zachariasse JM, Nieboer D, Oostenbrink R, Moll HA, Steyerberg EW. Multiple performance measures are needed to evaluate triage systems in the emergency department. *J Clin Epidemiol*. 2018;94:27-34.

128. Zachariasse JM, Kuiper JW, de Hoog M, Moll HA, van Veen M. Safety of the Manchester Triage System to Detect Critically Ill Children at the Emergency Department. *J Pediatr.* 2016;177:232-237 e231.

129. Fleming S, Gill P, Jones C, Taylor JA, Van den Bruel A, Heneghan C, et al. The Diagnostic Value of Capillary Refill Time for Detecting Serious Illness in Children: A Systematic Review and Meta-Analysis. *PLoS One*. 2015;10(9):e0138155.

130. Van den Bruel A, Haj-Hassan T, Thompson M, Buntinx F, Mant D, European Research Network on Recognising Serious Infection i. Diagnostic value of clinical features at presentation to identify serious infection in children in developed countries: a systematic review. *Lancet (London, England)*. 2010;375(9717):834-845.

131. Seiger N, Maconochie I, Oostenbrink R, Moll HA. Validity of different pediatric early warning scores in the emergency department. *Pediatrics*. 2013;132(4):e841-850.

132. Craig JC, Williams GJ, Jones M, Codarini M, Macaskill P, Hayen A, et al. The accuracy of clinical symptoms and signs for the diagnosis of serious bacterial infection in young febrile children: prospective cohort study of 15 781 febrile illnesses. *BMJ (Clinical research ed)*. 2010;340:c1594.

133. Fleming S, Thompson M, Stevens R, Heneghan C, Pluddemann A, Maconochie I, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *Lancet (London, England)*. 2011;377(9770):1011-1018.

134. Gravel J, Opatrny L, Gouin S. High rate of missing vital signs data at triage in a paediatric emergency department. *Paediatr Child Health*. 2006;11(4):211-215.

135. Bird C, Shea A, Michie CA, George G. A simple intervention improves the recording of vital signs in children presenting to the emergency department. *Emerg Med J*. 2009;26(10):698-700.

136. Benner P, Tanner C. Clinical judgment: how expert nurses use intuition. *Am J Nurs*. 1987;87(1):23-31.

137. Johansen ML, O'Brien JL. Decision Making in Nursing Practice: A Concept Analysis. *Nurs Forum*. 2016;51(1):40-48.

138. Edwards B. Walking in--initial visualisation and assessment at triage. *Accid Emerg Nurs*. 2007;15(2):73-78.

139. Van den Bruel A, Haj-Hassan T, Thompson M, Buntinx F, Mant D. Diagnostic value of clinical features at presentation to identify serious infection in children in developed countries: a systematic review. *Lancet (London, England)*. 2010;375(9717):834-845.

140. Parshuram CS, Duncan HP, Joffe AR, Farrell CA, Lacroix JR, Middaugh KL, et al. Multicentre validation of the bedside paediatric early warning system score: a severity of illness score to detect evolving critical illness in hospitalised children. *Critical care (London, England)*. 2011;15(4):R184.

141. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59(10):1087-1091.

142. Van den Bruel A, Thompson M, Buntinx F, Mant D. Clinicians' gut feeling about serious infections in children: observational study. *BMJ (Clinical research ed)*. 2012;345:e6144.

143. van der Vegt AE, Holman M, ter Maaten JC. The value of the clinical impression in recognizing and treating sepsis patients in the emergency department. *Eur J Emerg Med*. 2012;19(6):373-378.

144. Tibballs J, Kinney S, Duke T, Oakley E, Hennessy M. Reduction of paediatric in-patient cardiac arrest and death with a medical emergency team: preliminary results. *Arch Dis Child*. 2005;90(11):1148-1152.

145. Silverman MA, Walker AR, Nicolaou DD, Bono MJ. The frequency of blood pressure measurements in children in four EDs. *Am J Emerg Med*. 2000;18(7):784-788.

146. Marlais M, Lyttle MD, Inwald D. Ten concerns about blood pressure measurement and targets in paediatric sepsis. *Intensive Care Med*. 2017;43(3):433-435.

147. Sarganas G, Schaffrath Rosario A, Berger S, Neuhauser HK. An unambiguous definition of pediatric hypotension is still lacking: gaps between two percentile-based definitions and PALS / ATLS guidelines. *J Trauma Acute Care Surg*. 2018.

148. Brierley J, Carcillo JA, Choong K, Cornell T, Decaen A, Deymann A, et al. Clinical practice parameters for hemodynamic support of pediatric and neonatal septic shock: 2007 update from the American College of Critical Care Medicine. *Crit Care Med*. 2009;37(2):666-688.

149. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med*. 2013;41(2):580-637.

150. Rady MY, Smithline HA, Blake H, Nowak R, Rivers E. A comparison of the shock index and conventional vital signs to identify acute, critical illness in the emergency department. *Ann Emerg Med*. 1994;24(4):685-690.

151. Wira CR, Francis MW, Bhat S, Ehrman R, Conner D, Siegel M. The shock index as a predictor of vasopressor use in emergency department patients with severe sepsis. *West J Emerg Med*. 2014;15(1):60-66.

152. Balhara KS, Hsieh YH, Hamade B, Circh R, Kelen GD, Bayram JD. Clinical metrics in emergency medicine: the shock index and the probability of hospital admission and inpatient mortality. *Emerg Med J*. 2017;34(2):89-94.

153. Yasaka Y, Khemani RG, Markovitz BP. Is shock index associated with outcome in children with sepsis/septic shock?*. *Pediatr Crit Care Med*. 2013;14(8):e372-379.

154. Rousseaux J, Grandbastien B, Dorkenoo A, Lampin ME, Leteurtre S, Leclerc F. Prognostic value of shock index in children with septic shock. *Pediatr Emerg Care*. 2013;29(10):1055-1059.

155. Acker SN, Bredbeck B, Partrick DA, Kulungowski AM, Barnett CC, Bensard DD. Shock index, pediatric age-adjusted (SIPA) is more accurate than age-adjusted hypotension for trauma team activation. *Surgery*. 2017;161(3):803-807.

156. Linnaus ME, Notrica DM, Langlais CS, St Peter SD, Leys CM, Ostlie DJ, et al. Prospective validation of the shock index pediatric-adjusted (SIPA) in blunt liver and spleen trauma: An ATOMAC+ study. *J Pediatr Surg*. 2017;52(2):340-344.

157. Turner NM, Leroy PL. *Advanced paediatric life support*. 5th ed: Bohn Stafleu van Loghum; 2017.

158. Kleinman ME, Chameides L, Schexnayder SM, Samson RA, Hazinski MF, Atkins DL, et al. Pediatric Advanced Life Support: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *PEDIATRICS*. 2010;126(5):e1361-e1399.

159. Davis AL, Carcillo JA, Aneja RK, Deymann AJ, Lin JC, Nguyen TC, et al. American College of Critical Care Medicine Clinical Practice Parameters for Hemodynamic Support of Pediatric and Neonatal Septic Shock. *Critical Care Medicine*. 2017;45(6):1061-1093.

160. Parshuram CS, Dryden-Palmer K, Farrell C, Gottesman R, Gray M, Hutchison JS, et al. Effect of a Pediatric Early Warning System on All-Cause Mortality in Hospitalized Pediatric Patients: The EPOCH Randomized Clinical Trial. *Jama*. 2018;319(10):1002-1012.

161. Herbert R. Confidence Interval Calculator. https://www.pedro.org.au/english/downloads/confidence-interval-calculator/. Published 2013. Accessed 01-10-2017.

162. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet (London, England)*. 2005;365(9469):1500-1505.

163. Rappaport LD, Deakyne S, Carcillo JA, McFann K, Sills MR. Age- and sex-specific normal values for shock index in National Health and Nutrition Examination Survey 1999-2008 for ages 8 years and older. *Am J Emerg Med*. 2013;31(5):838-842.

164. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35.

165. Eytan D, Goodwin AJ, Greer R, Guerguerian AM, Laussen PC. Heart Rate and Blood Pressure Centile Curves and Distributions by Age of Hospitalized Critically Ill Children. *Front Pediatr*. 2017;5:52.

166. Abdelrazeq S, Ray S, Rogers L, Noren DP, Peters MJ, Inwald DP. Age-associated blood pressure distributions in paediatric intensive care units differ from healthy children. *Intensive Care Med*. 2017.

167. Wang EN, Chan C, Anand S, Dannenberg B. Considerations in the Management of Shock in the Pediatric Trauma Patient. *Pediatric Emergency Medicine Reports*. 2006;11(11).

168. Poor KM, Ducklow TB. Benefit of BP Measurement in Pediatric ED Patients. *ISRN Nursing*. 2012;2012:6.

169. Chaiyakulsil C, Pandee U. Validation of pediatric early warning score in pediatric emergency department. *Pediatr Int*. 2015;57(4):694-698.

170. Gold DL, Mihalov LK, Cohen DM. Evaluating the Pediatric Early Warning Score (PEWS) system for admitted patients in the pediatric emergency department. *Acad Emerg Med*. 2014;21(11):1249-1256.

171. Suttipongkaset P, Chaikittisilpa N, Vavilala MS, Lele AV, Watanitanon A, Chandee T, et al. Blood Pressure Thresholds and Mortality in Pediatric Traumatic Brain Injury. *Pediatrics*. 2018;142(2).

172. Hagedoorn NN, Zachariasse JM, Moll HA. Association between hypotension and serious illness in the emergency department: an observational study. *Arch Dis Child*. 2019.

173. Advanced Life Support Group. *Advanced Paediatric Life Support: The Practical Approach*. 5th ed: BMJ publishing group; 2011.

174. Goldstein B, Giroir B, Randolph A, International Consensus Conference on Pediatric S. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med*. 2005;6(1):2-8.

175. The World Bank Group. World Bank Country and Lending Groups. http://data.worldbank.org/about/country-and-lending-groups#Lower_middle_income Accessed 27-07-2015.

176. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol*. 2007;36(3):666-676.

177. Antal M, Regoly-Merei A, Nagy K, Greiner E, Biro L, Domonkos A, et al. Representative study for the evaluation of age- and gender-specific anthropometric parameters and blood pressure in an adolescent Hungarian population. *Ann Nutr Metab*. 2004;48(5):307-313.

178. Barba G, Buck C, Bammann K, Hadjigeorgiou C, Hebestreit A, Marild S, et al. Blood pressure reference values for European non-overweight school children: the IDEFICS study. *Int J Obes (Lond)*. 2014;38:S48-S56.

179. Blake KV, Gurrin LC, Evans SF, Newnham JP, Landau LI, Stanley FJ, et al. Reference ranges for blood pressure in preschool Australians, obtained by oscillometry. *J Paediatr Child Health*. 2000;36(1):41-46.

180. Grajda A, Kulaga Z, Gurzkowska B, Wojtylo M, Gozdz M, Litwin M. Preschool children blood pressure percentiles by age and height. *J Hum Hypertens*. 2017;31(6):400-408.

181. Hediger ML, Schall JI, Katz SH. Resting blood pressure and pulse rate distributions in black adolescents: The Philadelphia Blood Pressure Project. *PEDIATRICS*. 1984;74(6):1016-1021.

182. Kent AL, Kecskes Z, Shadbolt B, Falk MC. Blood pressure in the first year of life in healthy infants born at term. *Pediatric Nephrology*. 2007;22(10):1743-1749.

183. Karmar RT, Holtback U, Anita Bergh A, Svensson E, Wuhl E. Oscillometric casual blood pressure normative standards for Swedish children using ABPM to exclude casual hypertension. *Am J Hypertens*. 2014;28(4):459-468.

184. Krzyzaniak A, Krzywinska-Wiewiorowska M, Stawinska-Witoszynska B, Kaczmarek M, Krzych L, Kowalska M, et al. Blood pressure references for Polish children and adolescents. *Eur J Pediatr*. 2009;168(11):1335-1342.

185. Lurbe E, Redon J, Liao YL, Tacons J, Cooper RS, Alvarez V. Ambulatory Blood-Pressure Monitoring in Normotensive Children. *Journal of Hypertension*. 1994;12(12):1417-1423.

186. Rosner B, Cook N, Portman R, Daniels S, Falkner B. Determination of blood pressure percentiles in normal-weight children: some methodological issues. *Am J Epidemiol*. 2008;167(6):653-666.

187. National High Blood Pressure Education Program Working Group on High Blood Pressure in children and adolescents. The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. *PEDIATRICS*. 2004;114(2 Suppl 4th Report):555-576.

188. Satoh M, Inoue R, Tada H, Hosaka M, Metoki H, Asayama K, et al. Reference values and associated factors for Japanese newborns' blood pressure and pulse rate: the babies' and their parents' longitudinal observation in Suzuki Memorial Hospital on intrauterine period (BOSHI) study. *J Hypertens*. 2016;34(8):1578-1585.

189. Schwandt P, Scholze JE, Bertsch T, Liepold E, Haas GM. Blood pressure percentiles in 22,051 German children and adolescents: The PEP Family Heart Study. *Am J Hypertens*. 2015;28(5):672-679.

190. Weiss NS, Hamill PV, Drizd T. Blood Pressure levels of Children 6-11 Years: Relationship to Age, Sex, Race, and Socioeconomic Status, United States. *Vital Health Stat [11]*. 1973(135):1-30.

191. Barry P, Morris K, Ali T. *Paediatric Intensive Care*: OUP Oxford; 2010.

192. Rogers MC. *Textbook of Pediatric Intensive Care*: Williams & Wilkins; 1992.

193. Goldstein B, Giroir B, Randolph A. Reply: Values for Systolic Blood Pressure. *Pediatric Critical Care Medicine*. 2005;6(4):500-501.

194. Hoekelman RA. *Primary Pediatric Care*: Mosby; 1997.

195. Barkin RM. *Pediatric Emergency Medicine 2nd edition*: Mosby; 1996.

196. Nadas A. *Pediatric cardiology 3rd edition*: WB Saunders Co. ; 1976.

197. Behrman RE, Kliegman R, Jenson HB. *Nelson Textbook of Pediatrics*: Saunders; 2004.

198. de Caen AR, Berg MD, Chameides L, Gooden CK, Hickey RW, Scott HF, et al. Part 12: Pediatric Advanced Life Support: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*. 2015;132(18 Suppl 2):S526-542.

199. Spaite DW, Hu C, Bobrow BJ, Chikani V, Barnhart B, Gaither JB, et al. Association of Out-of-Hospital Hypotension Depth and Duration With Traumatic Brain Injury Mortality. *Ann Emerg Med*. 2017.

200. Asfar P, Meziani F, Hamel JF, Grelon F, Megarbane B, Anguel N, et al. High versus low blood-pressure target in patients with septic shock. *N Engl J Med*. 2014;370(17):1583-1593.

201. Lamontagne F, Day AG, Meade MO, Cook DJ, Guyatt GH, Hylands M, et al. Pooled analysis of higher versus lower blood pressure targets for vasopressor therapy septic and vasodilatory shock. *Intensive Care Med*. 2018;44(1):12-21.

202. Park MK, Menard SW, Yuan C. Comparison of auscultatory and oscillometric blood pressures. *Arch Pediatr Adolesc Med*. 2001;155(1):50-53.

203. Joffe R, Duff J, Garcia Guerra G, Pugh J, Joffe AR. The accuracy of blood pressure measured by arterial line and non-invasive cuff in critically ill children. *Critical care (London, England)*. 2016;20(1):177.

204. Ray S, Rogers L, Noren DP, Dhar R, Nadel S, Peters MJ, et al. Risk of over-diagnosis of hypotension in children: a comparative analysis of over 50,000 blood pressure measurements. *Intensive Care Med*. 2017;43(10):1540-1541.

205. Holt TR, Withington DE, Mitchell E. Which pressure to believe? A comparison of direct arterial with indirect blood pressure measurement techniques in the pediatric intensive care unit. *Pediatr Crit Care Med*. 2011;12(6):e391-394.

206. Institute Of Medicine. IOM report: the future of emergency care in the United States health system. *Acad Emerg Med*. 2006;13(10):1081-1085.

207. Roland D, Oliver A, Edwards ED, Mason BW, Powell CV. Use of paediatric early warning systems in Great Britain: has there been a change of practice in the last 7 years? *Arch Dis Child*. 2014;99(1):26-29.

208. Sandell JM, Maconochie IK. Paediatric early warning systems (PEWS) in the ED. *Emerg Med J*. 2016;33(11):754-755.

209. Roland D. Paediatric Early Warning Scores: Holy Grail and Achilles' heel. *Arch Dis Child Educ Pract Ed*. 2012;97(6):208-215.

210. Oldroyd C, Day A. The use of pediatric early warning scores in the emergency department. *J Emerg Nurs*. 2011;37(4):374-376; quiz 424.

211. Bradman K, Borland M, Pascoe E. Predicting patient disposition in a paediatric emergency department. *J Paediatr Child Health*. 2014;50(10):E39-44.

212. Nielsen KR, Migita R, Batra M, Gennaro JLD, Roberts JS, Weiss NS. Identifying High-Risk Children in the Emergency Department. *J Intensive Care Med*. 2016;31(10):660-666.

213. Lillitos PJ, Hadley G, Maconochie I. Can paediatric early warning scores (PEWS) be used to guide the need for hospital admission and predict significant illness in children presenting to the emergency department? An assessment of PEWS diagnostic accuracy using sensitivity and specificity. *Emerg Med J*. 2016;33(5):329-337.

214. Vredebregt SJ, Moll HA, Smit FJ, Verhoeven JJ. Recognizing critically ill children with a modified pediatric early warning score at the emergency department, a feasibility study. *Eur J Pediatr*. 2019;178(2):229-234.

215. Lillitos PJ, Maconochie IK. Paediatric early warning systems (PEWS and Trigger systems) for the hospitalised child: time to focus on the evidence. *Arch Dis Child*. 2017;102(6):479-480.

216. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Develpment, Validating, and Updating*: Springer; 2009.

217. Egdell P, Finlay L, Pedley DK. The PAWS score: validation of an early warning scoring system for the initial assessment of children in the emergency department. *Emerg Med J*. 2008;25(11):745-749.

218. Chapman SM, Wray J, Oulton K, Peters MJ. Systematic review of paediatric track and trigger systems for hospitalised children. *Resuscitation*. 2016;109:87-109.

219. National Institute for Health and Care Excellence (NICE). Fever in under 5s: assessment and initial management. 2013.

220. Nederlandse Vereniging voor Kindergeneeskunde (NVK). Richtlijn Koorts in de tweede lijn bij kinderen van 0 - 16 jaar. 2013.

221. Sambeeck SJV, Fuijkschot J, Kramer BW, Vos GD. Pediatric Early Warning System Scores: Lessons to be Learned. *J Pediatr Intensive Care*. 2018;7(1):27-32.

222. O'Mahony L, O'Mahony DS, Simon TD, Neff J, Klein EJ, Quan L. Medical complexity and pediatric emergency department and inpatient utilization. *Pediatrics*. 2013;131(2):e559-565.

223. Chan T, Rodean J, Richardson T, Farris RWD, Bratton SL, Di Gennaro JL, et al. Pediatric Critical Care Resource Use by Children with Medical Complexity. *J Pediatr.* 2016;177:197-203 e191.

224. Simon TD, Berry J, Feudtner C, Stone BL, Sheng X, Bratton SL, et al. Children with complex chronic conditions in inpatient hospital settings in the United States. *Pediatrics*. 2010;126(4):647-655.

225. Emergency preparedness for children with special health care needs. Committee on Pediatric Emergency Medicine. American Academy of Pediatrics. *Pediatrics*. 1999;104(4):e53.

226. Odetola FO, Gebremariam A, Davis MM. Comorbid illnesses among critically ill hospitalized children: Impact on hospital resource use and mortality, 1997-2006. *Pediatr Crit Care Med*. 2010;11(4):457-463.

227. Bhat N, Wright JG, Broder KR, Murray EL, Greenberg ME, Glover MJ, et al. Influenza-associated deaths among children in the United States, 2003-2004. *N Engl J Med*. 2005;353(24):2559-2567.

228. Sutton D, Stanley P, Babl FE, Phillips F. Preventing or accelerating emergency care for children with complex healthcare needs. *Arch Dis Child*. 2008;93(1):17-22.

229. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3:18.

230. Wuerz RC, Travers D, Gilboy N, Eitel DR, Rosenau A, Yazhari R. Implementation and refinement of the emergency severity index. *Acad Emerg Med*. 2001;8(2):170-176.

231. Farmer R, Mathur R, Bhaskaran K, Eastwood SV, Chaturvedi N, Smeeth L. Promises and pitfalls of electronic health record analysis. *Diabetologia*. 2018;61(6):1241-1248.

232. Kim E, Rubinstein SM, Nead KT, Wojcieszynski AP, Gabriel PE, Warner JL. The Evolving Use of Electronic Health Records (EHR) for Research. *Semin Radiat Oncol*. 2019;29(4):354-361.

233. Ross MK, Wei W, Ohno-Machado L. "Big data" and the electronic health record. *Yearb Med Inform*. 2014;9(1):97-104.

234. Kimia AA, Savova G, Landschaft A, Harper MB. An Introduction to Natural Language Processing: How You Can Get More From Those Electronic Notes You Are Generating. *Pediatr Emerg Care*. 2015;31(7):536-541.

235. Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet*. 2011;12(6):417-428.

236. Ambulatory Health Care Data: Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/ahcd/index.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fnchs%2Fahcd.htm. Accessed May 27, 2020.

237. Letsel Informatie Systeem (LIS). https://www.veiligheid.nl/organisatie/monitoring-onderzoek/letsel-informatie-systeem. Accessed July 1st, 2020.

238. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. *Ann Emerg Med*. 2018;71(5):565-574 e562.

239. Goto T, Camargo CA, Jr., Faridi MK, Freishtat RJ, Hasegawa K. Machine Learning-Based Prediction of Clinical Outcomes for Children During Emergency Department Triage. *JAMA Netw Open*. 2019;2(1):e186937.

240. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Jr., Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. *Critical care (London, England)*. 2019;23(1):64.

241. Kwon JM, Jeon KH, Lee M, Kim KH, Park J, Oh BH. Deep Learning Algorithm to Predict Need for Critical Care in Pediatric Emergency Departments. *Pediatr Emerg Care*. 2019.

242. Roquette BP, Nagano H, Marujo EC, Maiorano AC. Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. *Neural Netw*. 2020;126:170-177.

243. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22.

244. 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.

245. van Os HJA, Ramos LA, Hilbert A, van Leeuwen M, van Walderveen MAA, Kruyt ND, et al. Predicting Outcome of Endovascular Treatment for Acute Ischemic Stroke: Potential Value of Machine Learning Algorithms. *Front Neurol*. 2018;9:784.

246. Coggins SA, Weitkamp JH, Grunwald L, Stark AR, Reese J, Walsh W, et al. Heart rate characteristic index monitoring for bloodstream infection in an NICU: a 3-year experience. *Arch Dis Child Fetal Neonatal Ed*. 2016;101(4):F329-332.

247. Despins LA. Automated Deterioration Detection Using Electronic Medical Record Data in Intensive Care Unit Patients: A Systematic Review. *Comput Inform Nurs*. 2018;36(7):323-330.

248. Yanos PT, Ziedonis DM. The patient-oriented clinician-researcher: advantages and challenges of being a double agent. *Psychiatr Serv*. 2006;57(2):249-253.

249. Rosenberg L. Physician-scientists--endangered and essential. *Science*. 1999;283(5400):331-332.

250. Somekh I, Somekh E, Pettoello-Mantovani M, Somech R. The Clinician Scientist, a Distinct and Disappearing Entity. *J Pediatr*. 2019;212:252-253 e252.

II LIST OF ABBREVIATIONS

APLS	Advanced Paediatric Life Support
AUC	Area under the curve
AVPU	Alert, Verbal, Pain, Unresponsive
BP	Blood pressure
CI	Confidence interval
CRT	Capillary refill time
CTAS	Canadian Triage and Acuity Scale
DOR	Diagnostic odds ratio
ED	Emergency department
ENT	Ear, nose and throat
ESI	Emergency Severity Index
EUSEM	European Society for Emergency Medicine
FN	False negative
FP	False positive
GCS	Glasgow coma scale
GP	General practitioner
ICD-9	International classification of diseases, Ninth revision
ICU	Intensive care unit
IQR	Interquartile range
LR	Likelihood ratio
MAP	Mean arterial pressure
MTS	Manchester Triage System
OR	Odds ratio
PALS	Paediatric Advanced Life Support
PEWS	Paediatric early warning score
PMCA	Pediatric Medical Complexity Algorithm
PIM	Paediatric Index of Mortality
PRISM	Pediatric Risk of Mortality Score
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
TN	True negative
TP	True positive

III AUTHORS AND AFFILIATIONS

Claudio F Alves, MD

Hospital Professor Doutor Fernando da Fonseca, Department of Paediatrics, Emergency Unit, Amadora, Lisbon, Portugal

Paul Avillach, MD, PhD

Harvard Medical School, Department of Biomedical Informatics, Boston, MA, USA and: *Erasmus Medical Center, Department of Medical Informatics, Rotterdam, The Netherlands*

Dorine M Borensztajn, MD

Erasmus Medical Center-Sophia Children's Hospital, Department of Paediatrics, Rotterdam, the Netherlands

Pinky Rose Espina, MD

Medical University Vienna, Department of Pediatrics and Adolescent Medicine, Vienna, Austria

Paulo Freitas, MD, PhD

Hospital Professor Doutor Fernando da Fonseca, Intensive Care Unit, Amadora, Lisbon, Portugal

Susanne Greber-Platzer, MD, PhD

Medical University Vienna, Department of Pediatrics and Adolescent Medicine, Vienna, Austria

Nienke N Hagedoorn, MD

Erasmus Medical Center-Sophia Children's Hospital, Department of Paediatrics, Rotterdam, the Netherlands

Vera van der Hagen, MD

Erasmus Medical Center-Sophia Children's Hospital, Department of Paediatrics, Rotterdam, the Netherlands

Matthijs de Hoog, MD, PhD

Erasmus Medical Center-Sophia Children's Hospital, Department of Pediatric Intensive Care, Rotterdam, the Netherlands

Claudia MG Keyzer-Dekker, MD, PhD
Erasmus Medical Center-Sophia Children's Hospital, Department of Pediatric Surgery, Rotterdam, the Netherlands

Jan Willem Kuiper, MD, PhD
Erasmus Medical Center-Sophia Children's Hospital, Department of Pediatric Intensive Care, Rotterdam, the Netherlands

Dominique van der Lee, MD
Erasmus Medical Center-Sophia Children's Hospital, Department of Paediatrics, Rotterdam, the Netherlands
Currently: *Leiden University Medical Center, Department of Pediatrics, the Netherlands*

Johan van der Lei, MD, PhD
Erasmus University Medical Center, Department of Medical Informatics, Rotterdam, the Netherlands

Kevin Mackway-Jones, MD, PhD
Manchester Royal Infirmary, Emergency Department, Manchester, United Kingdom

Ian K Maconochie, MD, PhD
Imperial College NHS Healthcare Trust, Department of Paediatric Emergency Medicine, London, United Kingdom

Henriëtte A Moll, MD, PhD
Erasmus Medical Center-Sophia Children's Hospital, Department of Paediatrics, Rotterdam, the Netherlands

Daan Nieboer, MSc
Erasmus Medical Center, Department of Public Health, Rotterdam, The Netherlands

Ruud Nijman, MD, PhD
Imperial College NHS Healthcare Trust, Department of Paediatric Emergency Medicine, London, United Kingdom and *Imperial College London, Department of Infectious Diseases, London, United Kingdom*

Rianne Oostenbrink, MD, PhD
Erasmus Medical Center-Sophia Children's Hospital, Department of Paediatrics, Rotterdam, the Netherlands

Pleunie P Rood, MD, PhD
Erasmus Medical Center, Department of Emergency Medicine, Rotterdam, The Netherlands

Gert R Roukema, MD
Department of Surgery, Maasstad Hospital, Rotterdam, The Netherlands

Nienke Seiger, MD, PhD
Erasmus Medical Center-Sophia Children's Hospital, Department of Paediatrics, Rotterdam, the Netherlands

Frank J Smit, MD
Maasstad Hospital, Department of Paediatrics, Rotterdam, The Netherlands

Ewout Steyerberg, PhD
Leiden University Medical Center, Department of Medical Statistics and Bioinformatics, the Netherlands and: *Erasmus Medical Center, Department of Public Health, Rotterdam, The Netherlands*

Maria N Tsolia, MD, PhD
National and Kapodistrian University of Athens, Second Department of Paediatrics, Athens, Greece

Mirjam van Veen , MD, PhD
Erasmus Medical Center-Sophia Children's Hospital, Department of Paediatrics, Rotterdam, the Netherlands
Currently: *Groene Hart Ziekenhuis, Department of Paediatrics, Gouda, The Netherlands*

Evelien de Vos-Kerkhof, MD, PhD
Erasmus Medical Center-Sophia Children's Hospital, Department of Paediatrics, Rotterdam, the Netherlands
Currently: *Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands*

IV LIST OF PUBLICATIONS

This thesis:

Zachariasse JM, Espina PR, Borensztajn DM, Nieboer D, Maconochie IK, Steyerberg E, van der Lei J, Greber-Platzer S, Moll HA. *Improving triage for children with comorbidity using the ED-PEWS: prospective observational study.* [Submitted]

Zachariasse JM, Maconochie IK, Nijman R, Greber-Platzer S, Smit FJ, Nieboer D, van der Lei J, Alves CF, Moll HA. *Improving the prioritization of children at the emergency department: Updating the Manchester Triage System using vital signs*. PLoS One. 2021 Feb 9;16(2):e0246324

Zachariasse JM, Nieboer D, Maconochie IK, Smit FJ, Alves CF, Greber-Platzer S, Tsolia MN, Steyerberg EW, Avillach P, van der Lei J, Moll HA. *Development and validation of a Paediatric Early Warning Score for use in the emergency department: a multicentre study.* Lancet Child Adolesc Health. 2020 Aug;4(8):583-591.

Zachariasse JM, van der Hagen V, Seiger N, Mackway-Jones K, van Veen M, Moll HA. *Performance of triage systems in emergency care: a systematic review and meta-analysis*. BMJ Open. 2019 May 28;9(5):e026471

Hagedoorn NN, Zachariasse JM, Moll HA. *A comparison of clinical paediatric guidelines for hypotension with population-based lower centiles: a systematic review.* Crit Care. 2019 Nov 27;23(1):380.

Hagedoorn NN, Zachariasse JM, Moll HA. *Association between hypotension and serious illness in the emergency department: an observational study.* Arch Dis Child. 2020 Jun;105(6):545-551

Zachariasse JM, Nieboer D, Oostenbrink R, Moll HA, Steyerberg EW. *Multiple performance measures are needed to evaluate triage systems.* J Clin Epidemiol. 2018 Feb;94:27-34

Zachariasse JM, van der Lee D, Seiger N, de Vos-Kerkhof E, Oostenbrink R, Moll HA. *The role of nurses' clinical impression in the first assessment of children at the emergency department.* Arch Dis Child. 2017 Nov;102(11):1052-1056

Zachariasse JM, Seiger N, Rood PP, Alves CF, Freitas P, Smit FJ, Roukema GR, Moll HA. *Validity of the Manchester Triage System in emergency care: A prospective observational study.* PloS One. 2017 Feb 2;12(2):e0170811

Zachariasse JM, Kuiper JM, de Hoog M, Moll HA, van Veen M. *Safety of the Manchester Triage System to detect critically ill children at the emergency department.* J Pediatr. 2016 Oct;177:232-237

Other publications:

Hagedoorn NN, Zachariasse JM, Borensztajn D, Adriaansens E, von Both U, Carrol ED, Eleftheriou I, Emonts M, van der Flier M, de Groot R, Herberg J, Kohlmaier B, Lim E, Maconochie I, Martinon-Torres F, Nijman R, Pokorn M, Rivero Calle I, Tsolia M, Zavadska D, Zenz W, Levin M, Vermont C, Moll HA on behalf of PERFORM consortium. *Shock Index in the early assessment of febrile children at the Emergency Department: a prospective multicentre study.* [Submitted]

Tan CD, Hagedoorn NN, Dewez JE, Borensztajn DM, von Both U, Carrol ED, Eleftheriou I, Emonts M, van der Flier M, de Groot R, Herberg J, Kohlmaier B, Levin M, Lim E, Maconochie I, Martinon-Torres F, Nijman R, Pokorn M, Rivero Calle I, Strle F, Tsolia M, Vermont, CL, Yeung S, Zachariasse JM, Zenz W, Zavadska D, Moll HA on behalf of PERFORM consortium. *Rapid viral testing and antibiotic prescription in febrile children with respiratory symptoms visiting emergency departments in Europe.* [Submitted]

Nijman RG, Borensztajn DH, Zachariasse JM, Hajema C, Freitas P, Greber-Platzer S, Smit FJ, Alves C, van der Lei J, Steyerberg EW, Maconochie IK, Moll HA. *A clinical prediction model to identify children at risk for revisits with serious illness to the emergency department: a prospective multicentre observational study.* [Submitted]

Schinkelshoek G, Borensztajn DM, Zachariasse JM, Maconochie IK, Alves CF, Freitas P, Smit FJ, van der Lei J, Steyerberg EW, Greber-Platzer S, Moll HA. *Management of children visiting the emergency department during out-of-office hours: an observational study.* BMJ Paediatr Open. 2020 Sep 15;4(1):e000687

Zachariasse JM, Borensztajn DM, Nieboer D, Alves CF, Greber-Platzer S, Keyzer-Dekker CMG, Maconochie IK, Steyerberg EW, Smit FJ, Moll HA. *Sex-Specific Differences in Children Attending the Emergency Department: Prospective Observational Study.* BMJ Open. 2020 Sep 17;10(9):e035918

Zachariasse JJM, van Heurn LWE, Dompeling E, Been JV, Late-presenting Congenital Diaphragmatic Hernia. Arch Dis Child. 2011 Jul;96(7):687

V ABOUT THE AUTHOR

Joany Zachariasse was born on August 9th 1986 in Vlissingen, the Netherlands. In 2004 she completed high school at the Christelijke Scholengemeenschap Walcheren and started her medical training at Maastricht University. During her studies she participated in the International Track of Medicine and completed internships in Australia, the UK and the USA. Moreover, she interrupted her training for one year to study Portuguese and work on the implementation of a new curriculum at the Nursing Faculty of the Universidade Católica de Moçambique. After obtaining her Medical Degree in 2011, she worked for three months as a physician in the outpatient clinic of Haydom Lutheran Hospital in Tanzania. Subsequently she worked as a resident at the Paediatric Department of the Maasstad Hospital and the Paediatric Intensive Care Unit of the Erasmus MC- Sophia Children's Hospital.

In 2013 she started a PhD project on the recognition of critically ill children at the emergency department under supervision of Prof. H.A. Moll and Prof. J. van der Lei. The project was a combined effort of the Department of General Paediatrics, Erasmus MC-Sophia Children's Hospital and the Department of Medical Informatics of the Erasmus MC. In 2015 she obtained her Master of Science in Clinical Epidemiology at the Netherlands Institute for Health Sciences (NIHES). As part of her PhD, she spent 4 months working on novel methods to analyse observational data at the Department of Biomedical Informatics at Harvard Medical School in Boston under supervision of dr. Paul Avillach. For her research on the improvement of the Manchester Triage System, she was awarded the European Academy of Paediatrics Young Investigator award.

In January 2018, Joany started her paediatric residency at the Franciscus Gasthuis (supervisor Nico Hartwig) and the Erasmus MC- Sophia (supervisors Matthijs de Hoog and Andrica de Vries). In the meantime, she has worked as a postdoctoral researcher at the Department of General Paediatrics and has been involved in several studies on vital signs and vital sign-based scores with a focus on the use of electronic health record data. In the future, she intends to continue combining clinical work with research as a clinician-researcher.

VI PHD PORTFOLIO

Erasmus MC Department: General Paediatrics
Research School: Netherlands Institute for Health Sciences (NIHES)
PhD period: July 2013 – December 2017; February 2020 – July 2020
Promotor(s): Prof. dr. Henriëtte A. Moll and Prof. dr. Johan van der Lei

1. PHD TRAINING	YEAR	WORKLOAD (ECTS)
General academic skills		
BROK course ('Basiscursus Regelgeving en Organisatie voor Klinisch Onderzoekers')	2014 and 2020	1.0
Integrity in research	2014	0.3
TULIPS Grant writing and presenting weekend	2015	1.0
Research skills		
Internal research meetings, Department of General Paediatrics, Erasmus MC	2013-2017	2.0
Joint research meetings, Department of General Paediatrics, Medical Informatics and Center for Medical Decision Making, Erasmus MC	2013-2017	1.0
Master of Science clinical epidemiology	2013-2014	40.0
Core curriculum		
Study design		
Biostatistical methods I: Basic principles		
Clinical epidemiology		
Methodologic topics in epidemiologic research		
Courses for the quantitative researcher		
Biostatistical methods II: Classical regression models		
Elective courses		
Intervention research and clinical trials		
Prognosis research		
Repeated measurements		
Missing values in clinical research		
Advanced topics in decision-making in medicine		
Principles of Research in Medicine		
Clinical Decision Analysis		
Methods of Public Health Research		
Markers and prognostic research		
The practice of epidemiologic analysis		
Health economics		
Topics in meta-analysis		
Methods of health services research		
Cohort studies		
Logistic regression		
History of epidemiologic ideas		

Seminars and workshops		
Annual meeting of the Manchester Triage System International reference group	2013, 2017	1.0
TULIPS Young investigators day	2013, 2014, 2016	1.5
National conferences		
Congress Dutch society of Paediatrics [Poster presentation]	2013, 2018	1.8
Sophia Research Days [Poster presentation]	2014	0.9
Symposium "Pediatric emergency care: how research and practice interact", Erasmus MC-Sophia [Oral presentation]	2014	0.9
Scientific Day Maasstad Ziekenhuis [Oral presentation]	2015	0.9
International conferences		
European congres on Pediatric Resuscitation and Emergency Medicine (PREM) *Ghent [Poster presentation]*	2013	0.9
Congress of the European Academy of Paediatric Societies (EAPS) *Lyon [Oral presentation], Barcelona [Poster presentation], Paris [Oral presentation and poster presentation], Online [Oral presentation]*	2013-2014, 2018, 2020	2.7
European Congress on Emergency Medicine *Amsterdam [Oral presentation], Vienna [2x Oral presentation], Athens [Invited speaker and Oral presentation]*	2014, 2016-2017	2.7
International Triage Conference *Manchester [Oral presentation]*	2017	0.9

2. TEACHING	**YEAR**	**WORKLOAD (ECTS)**
Supervising Master students		
PEWS in children with comorbidity Ulas Dogan, medical student, Erasmus University Rotterdam	2014	1.5
Blood pressure in children: What's the value? Nienke Hagedoorn, medical student, Leiden University	2015	1.5
Gut feeling of nurses at children's emergency department Dominique van der Lee, medical student, Leiden University	2015	1.5
Validity of the Manchester Triage System in different settings: A systematic review and observational study Vera van der Hagen, medical student, Erasmus University Rotterdam	2016	1.5
Diversity and comorbidity in the pediatric emergency department Laura Konings, medical student, Erasmus University Rotterdam	2016	1.5
Comorbidity in children visiting the ED: A literature review and an observational study Arianne van Rijn, medical student, Erasmus University Rotterdam	2016	1.5
The association between temperature change after paracetamol administration and the severity of illness of febrile children in the emergency department Eveline Lutgert, medical student, Erasmus University Rotterdam	2020	1.5
Resource utilization among children with comorbidity at the Emergency Department Jessica Khyali, medical student, Leiden University	2020	1.5

VII DANKWOORD

Onderzoek doe je niet alleen en vele mensen hebben een belangrijke bijdrage geleverd aan dit proefschrift. Graag wil ik hen allen ontzettend bedanken! Een aantal van hen wil ik hieronder in het bijzonder noemen.

Allereerst een speciaal woord van dank aan mijn promotoren, prof. dr. Moll en prof. dr. Van der Lei.

Henriëtte, bedankt voor het vertrouwen dat je in me hebt gesteld om dit project, "jouw kindje", te gaan doen. Je was nauw betrokken bij het onderzoek, maar gaf me ook het zetje in de rug om met mijn eigen ideeën aan de slag te gaan. Dank ook voor alle kansen die je me hebt gegeven om me verder te ontwikkelen als onderzoeker en als persoon. Naast dat ik ontzettend veel van mijn promotietijd geleerd heb, heb ik er ook enorm van genoten.

Johan, dank voor je betrokkenheid bij dit proefschrift en je kritische bijdrage bij de verschillende manuscripten. Jij hebt me naar Boston gestuurd om me onder te dompelen in de medische informatica, wat voor mij een onvergetelijke ervaring was. Ik heb het erg gewaardeerd dat je deur altijd voor me openstond.

Prof. dr. Ewout Steyerberg en Ir. Daan Nieboer; dank voor jullie onmisbare bijdrage aan de methodologie in dit proefschrift. Ewout, dank voor je bijdrage aan de manuscripten en de waardevolle methodologische adviezen. Ik heb het erg gewaardeerd dat je voor mijn projecten, maar ook voor mij persoonlijk tijd wilde vrijmaken. Daarnaast ook veel dank voor het plaatsnemen in de leescommissie. Daan, bedankt voor je bijdrage aan de analyses van de verschillende studies. Of het nu ging om ingewikkelde R-error, of een methodologisch dilemma, dankzij jouw adviezen kon ik altijd weer vooruit.

Prof. dr. Matthijs de Hoog en Prof. dr. Edmond Rings: Dank dat jullie als leescommissie mijn manuscript hebben beoordeeld. Prof. Dr. Marjolein Berger en dr. Hester Lingsma, bedankt voor het plaatsnemen in de grote commissie. Prof. dr. Kevin Mackway-Jones, it is a great honour to have you on my committee.

I would like to thank all research partners for their participation in the TrIAGE study. Claudio Alves, Prof. dr. Susanne Greber-Platzer, dr. Ian Maconochie, and Frank Smit, without you this study would not have been possible. Questionnaires, discussions with ICT, ethical approval, site visits, reviews of manuscripts: thank you for your valuable time and contribution, and your belief in this research project.

Dorine Borensztajn, wat heb jij veel werk verzet aan de TrIAGE database. Hoewel het ons bloed, zweet en tranen heeft gekost, ben ik enorm trots op het resultaat!

Pinky Rose Espina, Ruud Nijman and Rikke Jorgensen, I am really grateful for your help with the data collection for this study and contribution to several of the studies.

Ook ben ik veel dank verschuldigd aan Idse Visser, Wouter van Rijsinge, Robert van Tilburg en Peter Keller, voor hun hulp bij de data-extractie van de verschillende projecten.

Paul Avillach, thank you very much for the warm welcome in the Avillach lab and the introduction to the world of biomedical informatics. I am very grateful for the opportunity to work with real "big data" and learn about all the possibilities when multiple sources of biomedical data are integrated. The great atmosphere at the department and your lab in particular has motivated me to "think big" in my own research as well. Hopefully, we can continue to collaborate in the future.

De onderzoeksgroep Paediatric Emergency Medicine en in het bijzonder Rianne Oostenbrink, Dorien Geurts en Dorine Borensztajn: dank voor jullie kritische vragen en inzichten bij de researchbesprekingen. Mijn voorgangers Mirjam van Veen en Nienke Seiger, bedankt voor jullie warm onthaal en de mooie artikelen die we samen hebben geschreven. Ook wil ik graag de (voormalig) masterstudenten Ulas Dogan, Dominique van der Lee, Arianne van Rijn, Laura Konings, Vera van der Hagen, Jessica Khyali en Eveline Lutgert bedanken, voor jullie enthousiasme tijdens het masteronderzoek en de belangrijke bijdrage aan de dataverzameling en een aantal van de artikelen.

Astrid en Saskia, bedankt voor jullie support en gezelligheid vanuit het secretariaat.

Een speciaal woord van dank voor de allerbeste kamergenoten. (Voormalige) SP-ers Yvette, Janske, Janneke, Thijs, Evelien, Myrthe, Iris, Nienke, Josephine, Maartje, Jesminne, Leontien, Britt, Doesjka, Chantal en Fabiënne, mijn onderzoeksperiode was niet half zo leuk geweest zonder jullie!

Lieve paranimphen Nienke en Josephine, wat ben ik blij dat jullie bij mijn verdediging naast mij staan! We hebben zoveel tijd op de SP samen doorgebracht, dat jullie inmiddels ook triage-experts zijn geworden. Niet alleen hebben we een hoop lief- en leed gedeeld maar bovenal was het met jullie altijd onwijs gezellig! Jullie gaan Rotterdam verlaten met mooie avonturen in het vooruitzicht, maar hopelijk kunnen we elkaar ook in de toekomst (digitaal) blijven zien.

Het laatste deel van mijn promotie heb ik gecombineerd met de opleiding tot kinderarts. Graag bedank ik mijn opleiders, Prof. dr. Matthijs de Hoog en dr. Andrica de Vries, Matthijs en Andrica, voor de mogelijkheid die jullie hebben geboden om kliniek en onderzoek te combineren. Dank ook aan de collega kinderartsen en arts-assistenten voor jullie interesse en support tijdens het afronden van mijn promotieonderzoek. Ik heb het geluk gehad om te mogen werken met vele fantastische dokters die voor mij een voorbeeld zijn geweest of nog steeds zijn op het gebied van kliniek, onderzoek of de

combinatie van beiden. Jasper, bedankt voor je luisterend oor en waardevolle adviezen als onderzoeksmentor.

Naast de vele mensen die een bijdrage hebben geleverd op professioneel vlak, ben ik heel dankbaar voor alle vrienden en familie die veel voor mij betekenen en die op hun eigen manier een bijdrage aan dit proefschrift hebben geleverd.

Allereerst heb ik enorm geluk met mijn lieve vriendinnetjes, uit Zeeland, Rotterdam en ver daarbuiten, voor jullie luisterend oor, gezelligheid, appjes, telefoontjes en bijkletsdates.

Ook wil ik graag mijn (schoon-)familie ontzettend bedanken voor alle betrokkenheid. Een paar speciale mensen wil ik hieronder nog in het bijzonder noemen.

Allereerst mijn schoonouders, Antonio en Gonnie, heel veel dank voor jullie belangstelling en support. Zonder alle oppas-uurtjes was dit proefschrift vast niet afgekomen en had ik op de donderdag nooit zo lekker kunnen eten. Graag wil ik jullie zeggen hoe erg ik het waardeer dat jullie altijd voor Bas en mij klaar staan.

Lieve tante Bep, zo'n buitengewone tante hebben er niet veel. Al van jongs af aan nam je me mee op sleeptouw, van het Middelburgse theater tot aan de gouden driehoek van India. Het is heel bijzonder hoe je niet alleen voor mij, maar ook voor vele anderen altijd klaar staat. Heel veel dank daarvoor!

Lieve paps, mams en Eline. Zoals in mijn laatste stelling verwoord, is jullie bijdrage aan dit proefschrift waarschijnlijk het grootst van allemaal. Dank voor alle kansen die jullie me hebben gegeven om uit te vliegen en het fijne thuis om altijd weer naar terug te keren.

Als allerlaatste, lieve Bas. Van Peru naar Japan en Boston en van Maastricht naar de Schieweg en het Molenlaankwartier. Wat hebben we de afgelopen jaren een hoop meegemaakt! Dankzij jouw liefde, support en zeker ook time-management is dit proefschrift nu echt af. Wat ben ik blij met jou en onze twee allerliefste mini's. Ik kijk uit naar de toekomst samen!