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

Summary

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 0·82–0·90) for high-urgency patients and 0·67 (0·61–0·73) 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 0·87–0·92) while a cutoff score of less than 6 was useful for identifying low-urgency patients with a sensitivity of 0·83 (0·81–0·85).

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

Funding Stichting de Drie Lichten, Stichting Sophia Kinderziekenhuis Fonds, and the European Union's Horizon 2020 research and innovation programme.

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.[1–4] 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.[5–7]

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.[8,9] 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.[10] 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.[11,12] The same scores, originally developed for the inpatient setting, are now being used in EDs.[13–16] 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.[9,11,17] A new PEWS designed for use in

Lancet Child Adolesc Health 2020; 4: 583–91

See **Comment** page 557

Department of General Paediatrics, Erasmus MC–Sophia Children's Hospital, University Medical Centre Rotterdam, Rotterdam, Netherlands (J M Zachariasse MD, Prof H A Moll PhD); **Department of Public Health** (D Nieboer MSc, Prof E W Steyerberg PhD) **and Department of Medical Informatics** (P Avillach PhD, Prof J van der Lei PhD), **Erasmus MC–University Medical Center Rotterdam, Rotterdam, Netherlands; Department of Paediatric Emergency Medicine, Imperial College NHS Healthcare Trust, London, UK** (I K Maconochie PhD); **Department of Paediatrics, Maasstad Hospital, Rotterdam, Netherlands** (F J Smit MD); **Department of Paediatrics, Emergency Unit, Hospital Professor Doutor Fernando da Fonseca, Lisbon, Portugal** (C F Alves MD); **Department of Pediatrics and Adolescent Medicine, Medical University Vienna, Vienna, Austria** (Prof S Greber-Platzer PhD); **National and Kapodistrian University of Athens, Second Department of Paediatrics, P and A Kyriakou Children's Hospital, Athens, Greece** (Prof M N Tsolia PhD); **Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands** (Prof E W Steyerberg); **and Harvard Medical School, Department of Biomedical Informatics, Boston, MA, USA** (P Avillach)

Correspondence to:
Prof Henriëtte A Moll, Department of General Paediatrics, Erasmus MC-Sophia Children's Hospital, University Medical Centre Rotterdam, PO Box 2060, 3000 CB, Rotterdam, Netherlands
h.a.moll@erasmusmc.nl

Research in context

Evidence before this study

We did a literature search to identify evidence on the use of Paediatric Early Warning Scores (PEWSs) in emergency care settings. We searched PubMed from inception to March 12, 2020, using the search terms "PEWS", "early warning", "risk assessment", "priority score", and "illness detection" in combination with terms related to the concept of "child" or "emergency department" or both. We defined PEWSs as scores, used for the assessment of a child's clinical status, based predominantly on physiological parameters, and applicable without computerised application (as opposed to triage systems or complex prediction models). We did not consider PEWSs designed for specific subgroups of children (eg, cardiac, neonatal). We identified four tools that fit our definition of an early warning score and were designed for use in the emergency department (ED): Egdell's Paediatric Advanced Warning Score, the Pediatric Assessment Triangle, the Pennine Acute Trust–Paediatric Observation Priority Score, and the Royal Manchester Children's Hospital Early Warning Score. Furthermore, we identified 12 reports validating PEWSs that were developed originally for use in hospitalised patients for the emergency care setting. These evaluations were mainly single-centre, retrospective evaluations, linking PEWSs with hospital or intensive care unit admission. None of the identified PEWSs were developed using statistical or data-driven methods. Given the absence of a single optimal score, and the use of unsatisfactory outcome measures, the value of PEWSs in the ED for identification of high-urgency and low-urgency patients is unclear.

Added value of this study

So far, no consensus exists regarding which PEWSs should be used in emergency care settings. The ED-PEWS is the first PEWS to be derived and validated fully on the basis of statistical methods and therefore aims to represent the optimal score based on currently available evidence. The ED-PEWS is based on a large cohort of more than 100 000 patients, consisting of the heterogeneous population of children visiting the ED. The multicentre and multinational nature of the study improves generalisability of the findings. The score consists of patient age and six physiological parameters that are commonly collected in the ED, is easy to calculate by hand, and is therefore practical to use in clinical practice.

Implications of all the available evidence

Available research shows that only a small proportion of children visiting the ED present with illness that requires immediate care, in our study 0·6–7·2%. Besides existing strategies, such as triage, novel methods have been sought to improve the prioritisation of children in the ED. Our results emphasise that the combination of multiple physiological parameters, more than a single parameter alone, is a strong predictor of patient urgency. The use of a single, optimal PEWS in the ED might assist health-care workers to recognise the small group of seriously ill patients, but also to distinguish the large subgroup of low-urgency patients. Further independent validation is needed to establish the value of the ED-PEWS in other settings—eg,, in low-resource settings and in different subgroups of patients.

See Online for appendix

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 exists 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 under age 16 years from five diverse EDs in four European countries: Erasmus MC (Rotterdam, Netherlands), Maasstad Hospital (Rotterdam, 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 (appendix p 2). The five participating study sites were diverse regarding type of hospital, number of ED visits, and complexity of the patient population (appendix p 2). All hospitals used an electronic hospital information system in which ED nurses routinely entered the collected clinical data. Triage was done routinely with the Manchester Triage System (MTS).[18] 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.[19] 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,[20] and because it was not a standard measurement in the study sites. At the participating EDs, physiological parameters were measured at the discretion of the nurse, according to local practices (appendix pp 3–4). 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.[21] Missing physiological measurements were imputed 25 times using the MICE algorithm in R (version 3.6.3; appendix pp 5–7). We assumed these items to be missing at random, which means that missingness can be fully accounted for by other variables in the database,[22] 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 analyses were done and pooled for a final result.[22]

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.[23]

Outcomes

As part of the TrIAGE project, a reference standard was developed, serving as a proxy for each child's true urgency (appendix p 8). This reference standard was developed based on the methodology of a previously published study, based on information from the entire ED visit.[24,25] 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.[7]

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.

	Erasmus MC (n=18 594)	Maasstad Hospital (n=10 584)	St Mary's Hospital (n=15 556)	Hospital Fernando da Fonseca (n=53 175)	Vienna General Hospital (n=21 300)
Sex					
Male	10 774 (57·9%)	6004 (56·7%)	8677 (55·8%)	27 685 (52·1%)	11 233 (52·7%)
Female	7820 (42·1%)	4580 (43·3%)	6879 (44·2%)	25 490 (47·9%)	10 067 (47·3%)
Age, years					
<1	3680 (19·8%)	1639 (15·5%)	2773 (17·8%)	7090 (13·3%)	3441 (16·2%)
1 to <12	11 855 (63·8%)	6556 (61·9%)	10 742 (69·1%)	38 447 (72·3%)	15 239 (71·5%)
≥12	3059 (16·5%)	2389 (22·6%)	2041 (13·1%)	7638 (14·4%)	2620 (12·3%)
MTS urgency					
Emergent or very urgent	2427 (13·1%)	1515 (14·3%)	1605 (10·3%)	6222 (11·7%)	1084 (5·1%)
Urgent	8745 (47·0%)	5110 (48·3%)	3961 (25·5%)	10 951 (20·6%)	3851 (18·1%)
Standard or non-urgent	6852 (36·9%)	3857 (36·4%)	9990 (64·2%)	36 002 (67·7%)	15 314 (71·9%)
Unknown	570 (3·1%)	102 (1·0%)	0	0	1051 (4·9%)
Diagnostics					
Laboratory	6234 (33·5%)	2102 (19·9%)	1558 (10·0%)	6990 (13·1%)	7841 (36·8%)
Imaging	4487 (24·1%)	3907 (36·9%)	2256 (14·5%)	12 624 (23·7%)	1677 (7·9%)
Medication					
Inhalation	762 (4·1%)	612 (5·8%)	1073 (6·9%)	5223 (9·8%)	951 (4·5%)
Intravenous	2182 (11·7%)	1360 (12·8%)	703 (4·5%)	4009 (7·5%)	989 (4·6%)
Disposition					
ICU admission or death at ED	520 (2·8%)	17 (0·2%)	26 (0·2%)	136 (0·3%)	15 (0·1%)
Hospital admission	3801 (20·4%)	2463 (23·3%)	1599 (10·3%)	2612 (4·9%)	1279 (6·0%)
Discharge or other	14 273 (76·8%)	8104 (76·6%)	13 931 (89·6%)	50 427 (94·8%)	20 006 (93·9%)

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

Table 1: **Demographics and baseline characteristics**

Decision curve analysis is a method of evaluating the performance of prediction models, taking into account their clinical consequences.[26] 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.[19,27–29] The Paediatric Advanced Warning Score by Egdell and colleagues[27] 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.[27] The Bedside PEWS by Parshuram and colleagues[28] 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.[29] 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 (appendix pp 2–3). 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 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 (appendix p 2). 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 0·6–7·2) and 29 127 (24·4%) of intermediate urgency (range across hospitals 20·2–39·6). 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 (p 9). 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 (appendix p 5).

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 (7·07, 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 7854 (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).

The cross-validated c-statistic of the ED-PEWS was 0·86 (95% prediction interval 0·82–0·90) for the identification of high-urgency patients and 0·67 (0·61–0·73) for the identification of high-urgency or intermediate-urgency

	Outcome		Odds ratio (95% CI)	
	High urgency	Intermediate urgency	Univariable	Multivariable
Hospital				
Erasmus MC	1335/18 594 (7·2%)	5940/18 594 (31·9%)	Reference	Reference
Maasstad Hospital	129/10 584 (1·2%)	4193/10 584 (39·6%)	0·97 (0·92–1·02)	0·77 (0·72–0·81)
St Mary's Hospital	275/15 556 (1·8%)	3211/15 556 (20·6%)	0·43 (0·41–0·45)	0·51 (0·48–0·54)
Hospital Fernando da Fonseca	1163/53 174 (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%)	4212/21 300 (19·8%)	0·37 (0·36–0·39)	0·45 (0·43–0·47)
Age, years				
<1	772/18 624 (4·1%)	5196/18 624 (27·9%)	Reference	Reference
1 to <2	454/15 230 (3·0%)	3614/15 230 (23·7%)	0·77 (0·73–0·81)	0·93 (0·88–0·99)
2 to <5	688/30 324 (2·3%)	6532/30 324 (21·5%)	0·66 (0·63–0·68)	1·10 (1·04–1·16)
5 to <12	748/37 284 (2·0%)	8374/37 284 (22·5%)	0·68 (0·65–0·70)	1·66 (1·56–1·78)
≥12	367/17 747 (2·1%)	4558/17 747 (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	2545/117 671 (2·2%)	27 617/117 671 (23·5%)	Reference	Reference
Decreased	485/1539 (31·5%)	657/1539 (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	1637/110 221 (1·5%)	23 024/110 221 (20·9%)	Reference	Reference
Mild-to-moderate increase	1120/8126 (13·8%)	4794/8126 (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	2820/117 569 (2·4%)	27 638/117 569 (23·5%)	Reference	Reference
Abnormal	210/1641 (12·8%)	636/1641 (38·8%)	3·42 (3·03–3·85)	1·70 (1·49–1·95)

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

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

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 (appendix p 10). Regarding our secondary outcome measures, the cross-validated c-statistic was 0·83 (95% prediction interval 0·77–0·89) for ICU admission and 0·69 (0·64–0·73) for hospital admission (table 3; appendix p 11).

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; appendix p 12).

We observed substantial heterogeneity in the predicted risk of high or low urgency between the different settings.

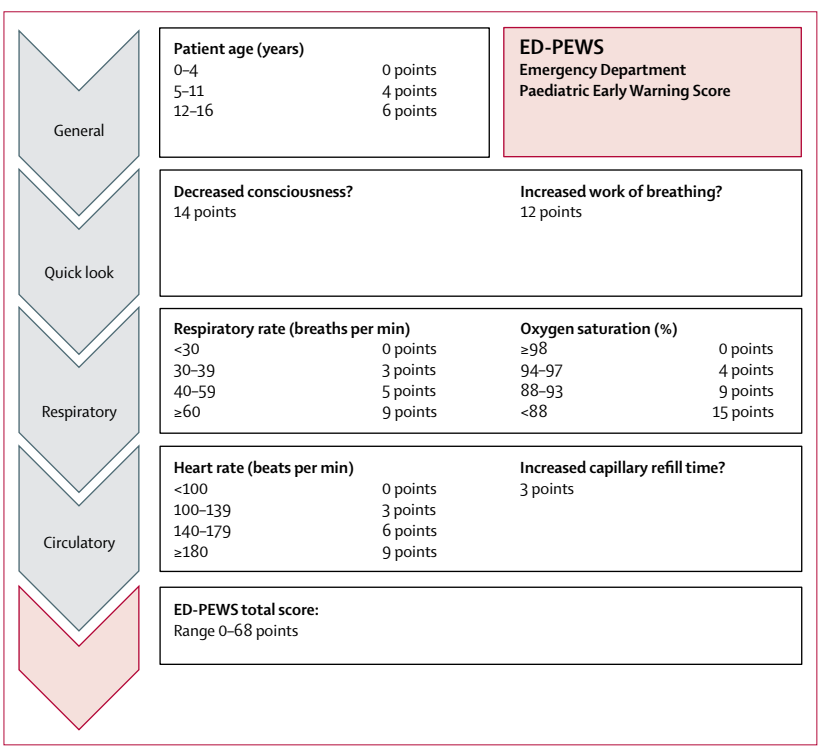

Figure: **The ED-PEWS**
ED-PEWS=Emergency Department Paediatric Early Warning Score.

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 (appendix p 13).

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

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 4542 febrile children, the c-statistic of the ED-PEWS for recognition of high-urgency patients was 0·86 (95% CI 0·81–0·91) for high-urgency patients and 0·62 (0·60–0·64) for intermediate-urgency patients (appendix p 14). The c-statistics for the secondary outcomes were 0·95 (95% CI 0·89–1·00) for ICU admission and 0·61 (0·58–0·63) for hospital admission. Specificity for the high-urgency cutoff was 0·95 (95% CI 0·90–1·00), compared with 0·90 (0·87–0·92) in the original cohort, and sensitivity for the low-urgency cutoff was 0·82 (95% CI 0·80–0·84) compared with 0·83 (0·81–0·85) in the original cohort (appendix p 14). We repeated the model development process in a complete case dataset. Coefficients in the complete case analysis were largely similar and did not

	Reference standard: high urgency *vs* intermediate or low urgency			Reference standard: high or intermediate urgency *vs* low urgency			ICU admission			Hospital admission		
	c-statistic	95% CI	95% prediction interval	c-statistic	95% CI	95% prediction interval	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	0·83	0·79–0·87	0·77–0·89	0·69	0·67–0·71	0·64–0·73
Parshuram PEWS[28]*	0·82	0·79–0·84	0·76–0·87	0·64	0·61–0·66	0·57–0·70	0·79	0·71–0·87	0·62–0·97	0·65	0·62–0·67	0·60–0·69
Egdell PEWS[27]†	0·81	0·80–0·83	0·78–0·84	0·63	0·61–0·65	0·59–0·68	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); categorise 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; categorise alert, verbal, pain, unresponsive scale into three categories instead of four.

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

lead to changes in variable selection (appendix p 15). Moreover, the apparent performance in the development population was lower for the model developed in the original (imputed) database.

In the appendix (p 16), 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 0·82–0·89) for the identification of high-urgency patients and 0·69 (0·63–0·74) for high-to-intermediate-urgency patients.

	Proportion of patients classified (range over hospitals)	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Cutoff 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)
Cutoff score for low urgency					
<6	27·9% (18·4–32·4)	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	47·2% (39·3–52·0)	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	59·4% (49·3–64·6)	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	63·9% (54·3–69·0)	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	71·1% (62·7–75·2)	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.

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

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.[21] For example, we handled missing data by multiple imputation, avoided stepwise selection methods, and maintained continuous physiological parameters as continuous. Thereby, we aimed for 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.[10,17,19,30] Only one published PEWS used the discriminative ability of the candidate parameters in the predictor selection process,[28] but combined the results with clinical judgment and used expert opinion to identify cutoff points for the different items. Moreover, most PEWSs have been developed for use in hospitalised children.[10,19] 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.[31–33] 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.[7,34] Therefore, we developed a reference standard (high, intermediate, and low urgency), based on literature and expert opinion, as a proxy for true patient 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 dichotomisation.

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,[20] and blood pressure is not recommended as an initial screening tool in the ED by several practice guidelines.[35,36]

The newly developed ED-PEWS consists of a combination of physiological measurements that can be easily and objectively obtained in 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 cutoff of at least 15, high-urgency patients can be identified with a high specificity of 0·90 (95% CI 0·87–0·92). 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 0·81–0·85). 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.[37] 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.

Contributors

All authors contributed to the conception and design of the study and interpretation of the findings. JMZ and DN designed and did the analyses. JMZ wrote the first draft of the manuscript. All authors revised it and gave their approval of the final version. HAM is the guarantor.

Declaration of interests

We declare no competing interests.

Acknowledgments

This study was supported by Stichting de Drie Lichten and Stichting Sophia Kinderziekenhuis Fonds. For data collection on the cohort of febrile children from the P and A Kyriakou Children's Hospital, funding was provided by the European Union's Horizon 2020 research and innovation programme under grant agreement number 668303. We thank the ED nursing staff of the participating hospitals for their participation and collection of data. We thank Nienke Seiger, Rikke Jorgensen, and Pinky Rose Espina for their assistance with data collection, and Dorine Borensztajn for her contribution to the development of the TrIAGE database.

References

1 Hostetler MA, Mace S, Brown K, et al. Emergency department overcrowding and children. *Pediatr Emerg Care* 2007; **23:** 507–15.
2 Wolfe I, Thompson M, Gill P, et al. Health services for children in western Europe. *Lancet* 2013; **381:** 1224–34.
3 Institute Of Medicine. IOM report: the future of emergency care in the United States health system. *Acad Emerg Med* 2006; **13:** 1081–85.
4 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:** 878–88.
5 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:** e1602–08.
6 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 Pediatrics* 2016; **177:** 232–37.e1.

7 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:** e026471.

8 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:** 26–29.

9 Sandell JM, Maconochie IK. Paediatric Early Warning Systems (PEWS) in the ED. *Emerg Med J* 2016; **33:** 754–55.

10 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:** e014497.

11 Roland D. Paediatric Early Warning Scores: Holy Grail and Achilles' heel. *Arch Dis Child Educ Pract Ed* 2012; **97:** 208–15.

12 Oldroyd C, Day A. The use of Pediatric Early Warning Scores in the emergency department. *J Emerg Nurs* 2011; **37:** 374–76.

13 Bradman K, Borland M, Pascoe E. Predicting patient disposition in a paediatric emergency department. *J Paediatr Child Health* 2014; **50:** E39–44.

14 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:** 660–66.

15 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:** 329–37.

16 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:** 229–34.

17 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:** 479–80.

18 Mackway-Jones K, Marsden J, Windle J. Emergency triage. UK: John Wiley & Sons, 2014.

19 Seiger N, Maconochie I, Oostenbrink R, Moll HA. Validity of different Pediatric Early Warning Scores in the emergency department. *Pediatrics* 2013; **132:** e841–50.

20 Hagedoorn NN, Zachariasse JM, Moll HA. Association between hypotension and serious illness in the emergency department: an observational study. *Arch Dis Child* 2019; **105:** 545–51.

21 Steyerberg EW. Clinical prediction models: a practical approach to development, validating, and updating. New York: Springer, 2009.

22 Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006; **59:** 1087–91.

23 Turner N, Leroy P. Advanced paediatric life support. Amsterdam: Reed Business, 2011.

24 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:** e0170811.

25 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 Technol Assess* 2007; **11:** iii, ix-51.

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

27 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:** 745–49.

28 Parshuram CS, Duncan HP, Joffe AR, 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. *Crit Care* 2011; **15:** R184.

29 Parshuram CS, Dryden-Palmer K, Farrell C, 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:** 1002–12.

30 Chapman SM, Wray J, Oulton K, Peters MJ. Systematic review of paediatric track and trigger systems for hospitalised children. *Resuscitation* 2016; **109:** 87–109.

31 van de Maat J, Jonkman H, van de Voort E, 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; **179:** 1097–106.

32 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:** 698–700.

33 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:** 211–15.

34 Seiger N, Moll HA. Triage systems: outcome measures to validate. *Ann Emerg Med* 2013; **61:** 372–73.

35 National Institute for Health and Care Excellence. Fever in under 5s: assessment and initial management. 2013. https://www.nice.org.uk/guidance/ng143 (accessed April 4, 2020).

36 Nederlandse Vereniging voor Kindergeneeskunde. Richtlijn Koorts in de tweede lijn bij kinderen van 0 - 16 jaar. 2013. https://werkboeken.nvk.nl/Kwaliteit/Richtlijnen-overzicht/Details/articleType/ArticleView/articleId/887 (accessed April 4, 2020)

37 Sambeeck SJV, Fuijkschot J, Kramer BW, Vos GD. Pediatric Early Warning System scores: lessons to be learned. *J Pediatr Intensive Care* 2018; **7:** 27–32.