Performance of Pediatric Mortality Prediction Models in Low- and Middle-Income Countries: A Systematic Review and Meta-Analysis

Fiona Muttalib, MD[1,2], Virginie Clavel, MD, MPH[3], Lauren H. Yaeger, MA, MLIS[4], Vibhuti Shah, MD[1,5], and Neill K. J. Adhikari, MD, MSc[1,6,7]

Objective To describe the performance of prognostic models for mortality or clinical deterioration events among hospitalized children developed or validated in low- and middle-income countries.

Study design A medical librarian systematically searched EMBASE, Ovid Medline, Scopus, Cochrane Library, EBSCO Global Health, LILACS, African Index Medicus, African Journals Online, African Healthline, Med-Carib, and Global Index Medicus (from 2000 to October 2019). We included citations that described the development or validation of a pediatric prognostic model for hospital mortality or clinical deterioration events in low- and middle-income countries. In duplicate and independently, we extracted data on included populations and model prognostic performance and evaluated risk of bias using the Prediction model Risk Of Bias Assessment Tool.

Results Of 41 279 unique citations, we included 15 studies describing 15 prognostic models for mortality and 3 models for clinical deterioration events. Six models were validated in >1 external cohort. The Lambarene Organ Dysfunction Score (0.85 [0.77-0.92]) and Signs of Inflammation in Children that Kill (0.85 [0.82-0.88]) had the highest summary C-statistics (95% CI) for discrimination. Calibration and classification measures were poorly reported. All models were at high risk of bias owing to inappropriate selection of predictor variables and handling of missing data and incomplete performance measure reporting.

Conclusions Several prognostic models for mortality and clinical deterioration events have been validated in single cohorts, with good discrimination. Rigorous validation that conforms to current standards for prediction model studies and updating of existing models are needed before clinical implementation. *(J Pediatr 2020;225:182-92)*.

Child mortality has decreased by >50% globally in the last 30 years, but 5.5 million children <5 years of age died in 2017 worldwide.[1] The importance of addressing deficiencies in hospital care management has been emphasized.[2-4] In low- and middle-income countries (LMICs), child hospital mortality remains high, ranging from 2.3% to 15% compared with <1% in high-income countries.[5-10] A significant proportion of this mortality (33%-85%) occurs within 48 hours of hospital admission.[11-13] In these settings, prediction models that enhance early identification of the sickest children are needed to guide timely referral and transport of patients, efficient allocation of resources, and counselling regarding anticipated clinical trajectories.[14,15]

Prediction models have been developed to identify children at greatest risk of in-hospital mortality or clinical deterioration in LMICs; however, none is routinely used in clinical practice or research.[12,15,16] Only a few scores have been externally validated and compared in large prospective cohorts.[6,15,17] A scoping review found limited evidence regarding the validity, reliability, and impact of pediatric early warning scores in resource-limited settings, but did not focus on models developed or validated in LMICs.[18] To date, no systematic review has summarized the evidence for performance of pediatric prognostic models developed or validated in LMICs. A lack of compelling evidence for a given prognostic model's performance may hinder widespread dissemination and adoption in clinical practice or research.

We undertook this systematic review to synthesize the evidence regarding the performance of pediatric prognostic models developed or validated in LMICs for prediction of hospital mortality or clinical deterioration events. The population of interest was acutely ill children (<18 years old) presenting to hospital in LMICs. The primary outcome was discrimination of prognostic models for

Bedside PEWS	Bedside Pediatric Early Warning Score
LMIC	Low- and middle-income countries
LODS	Lambarene Organ Dysfunction Score
PEDIA	Pediatric Early Death Index in Africa
PEWS	Pediatric Early Warning Score
PICU	pediatric intensive care unit
PRISM	Pediatric Risk of Mortality
PROBAST	Prediction model Risk Of Bias Assessment Tool
SICK	Signs of Inflammation that Kill

From the [1]Institute of Health Policy, Management and Evaluation, University of Toronto, [2]Center for Global Child Health, Hospital for Sick Children, Toronto, Ontario, and [3]Faculty of Medicine, Department of Pediatrics, McGill University, Montreal, Quebec, Canada; [4]Becker Medical Library Washington University School of Medicine in St. Louis, St. Louis, MO; [5]Department of Pediatrics, Mount Sinai Hospital, [6]Department of Critical Care Medicine, Sunnybrook Health Sciences Centre, and [7]Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Ontario, Canada

The authors declare no conflicts of interest.

Portions of this study were presented at the Canadian Critical Care Forum, November 11, 2019, Toronto, Canada.

https://doi.org/10.1016/j.jpeds.2020.05.016

mortality at hospital discharge and the secondary outcome was discrimination for clinical deterioration events.

Methods

Additional details are provided in **Appendix 1** (available at www.jpeds.com). The protocol for this systematic review and meta-analysis was registered in PROSPERO (CRD 42019125967), which was updated on June 7, 2019, to modify the outcome from prediction of 72-hour mortality to hospital mortality to capture prognostic model studies that did not restrict the timing of mortality determination.

Search Strategy

We searched the published literature using strategies created by a medical librarian for the concepts of LMICs, risk assessment, early warning, severity of illness, alert criteria, and children. The search strategy was implemented in EMBASE, Ovid Medline, Scopus, Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Cumulative Index of Nursing and Allied Health Literature, and EBSCO Global Health. Additional keyword searches were conducted in LILACS, African Index Medicus, African Journals Online, African Healthline, Med-Carib, and Global index medicus. We did not use a search filter for age because of concerns of imperfect sensitivity. Searches were completed in March 2019 and updated on October 29, 2019, with results limited to articles published from 2000 to reflect current trends in childhood mortality.[1] No language limits were applied. Fully reportable searches are listed in **Appendix 2**. Key article reference lists and reviews were hand searched for additional citations.

Study Selection

Studies were eligible for inclusion if they included children 0-18 years presenting or admitted to hospital in a LMIC with acute illness, reported the development or validation of a prognostic model with or without external validation, and reported model performance with respect to hospital mortality at any time point or occurrence of a clinical deterioration event.[19] We considered any study design (cohort, case series, quasiexperimental [eg, before–after], case control, or randomized trial) that met other inclusion criteria. We excluded studies that were unpublished or reported intensive care unit-specific prediction models, disease-specific prediction models, models developed or validated only in high-income countries, and models with no disaggregated pediatric outcome data. To obtain full-text articles, we searched the holdings of 3 university libraries and attempted to contact study authors.

Clinical deterioration events were defined within each study and included any of transfer to a referral hospital, high-dependency unit or intensive care unit, unexpected cardiac or respiratory arrest, provision of cardiopulmonary resuscitation, tracheal intubation, or administration of vasoactive medication.[8,20]

Data Management

References were downloaded into reference management software (EndNote, Clarivate Analytics, Philadelphia, Pennsylvania) and uploaded to Covidence (Veritas Health Innovation Ltd, Melbourne, Australia). Two independent reviewers screened titles and abstracts and examined the full text of any potentially relevant citation for inclusion. At all phases of the review, disagreement was resolved by consensus and adjudicated by a third reviewer.

Data Extraction and Risk of Bias Assessment

Two reviewers extracted data independently onto electronic case record forms (Microsoft Excel, Microsoft, Redmond, Washington). Where data were missing or unclear, study authors were contacted. Data elements were extracted using the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies checklist.[21]

The Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to assess risk of bias at the individual study level (n = 22) in 4 domains: (1) participant selection, (2) predictor selection and measurement, (3) outcome definition and determination, and (4) statistical analysis.[22] A high risk of bias was considered to downgrade the quality of evidence.

Summary Measures

Model calibration, discrimination, and classification measures for hospital mortality and clinical deterioration events were compared qualitatively across scores. An area under the receiver operator curve (C statistic) of <0.6 is considered poor discrimination, 0.6-0.75 possibly useful discrimination and, >0.75 clearly helpful discrimination.[23,24] Classification measures are summarized by sensitivity, specificity, predictive values and likelihood ratios with respect to the outcome of interest. Calibration is represented by outcome and events tables or a calibration plot, with perfect calibration represented by an observed to expected ratio of 1.[23]

Analysis

A meta-analysis was performed for discrimination of externally validated models using the random effects model, which was chosen with the assumption that variation in predictive performance may be due to heterogeneity between studies rather than chance alone.[23] A summary C-statistic, CI, and summary classification measures were calculated using the metamisc package in R studio (R studio version 1.2.1335, The R Foundation, Vienna, Austria).[25] Statistical heterogeneity was described using the I^2 measure, with values of >75%, 25%-75%, and <25% considered to reflect high, moderate, and low heterogeneity, respectively.[26] Calibration measures could not be analyzed across studies as observed and expected events were not reported in sufficient studies. We planned to assess publication bias if >10 studies examined a given prognostic model.[27]

Risk of Bias Across Studies

An adapted GRADE framework for evaluation of prognostic studies was used to judge the quality of evidence

across studies for validated prognostic models.[28,29] Factors that increased the quality of evidence included excellent discrimination and calibration and a dose-response gradient between an increase in prognostic model score and risk of death. Factors that decreased the quality of evidence included early phase of investigation, study limitations, high risk of bias, inconsistency, indirectness, and imprecision.

Results

The search yielded 41 279 unique citations (**Figure**; available at www.jpeds.com). We completed a full-text review of 374 studies and included 15 studies describing 15 discrete prognostic models for risk of mortality and 3 discrete models for clinical deterioration events.[5,6,12,15-17,30-38] We contacted 6 study authors for additional information, of whom 2 provided additional data.[15,17] Five studies described prognostic model development only, 3 described development and external validation, and 7 described validation only.[5,6,12,15-17,30-38] Seven prognostic models for mortality from 5 studies were validated in >1 external cohort and 6 of these were included in meta-analysis.[12,15,17,30,32] We calculated pooled measures for 1 prognostic model for mortality for which complete classification measures were reported in >1 cohort for the same threshold score.[12,17,32] None of the prognostic models for clinical deterioration events was validated in >1 cohort.

Study Characteristics

The development and validation study characteristics are summarized in **Table I**. Studies were conducted in sub-Saharan Africa (n = 9), South Asia (n = 3), Southeast Asia (n = 2), and Central America (n = 1).[5,6,12,15-17,30-38] Studies were primarily conducted in urban, tertiary hospitals (n = 14) and 3 included rural district health centers.[5,6,12,15-17,30-38] Most were prospective cohort studies of children with acute illness presenting to an emergency department or admitted to hospital (n = 9), and 6 studies had more restrictive inclusion criteria.[5,12,15,16,30-38] In most cases, predictor variables were clinical signs observable or measurable in the intended setting with minimal equipment.

The number of candidate variables (4-25) and the method for variable selection (stepwise backwards elimination, complete model analysis, univariable analysis, expert consensus) was variable. The event per variable rate in development studies ranged from 2.7 to 72.5 and the number of events in validation studies ranged from 5 to 556. Final prediction models comprised elements of the clinical history and characteristics of the neurologic, cardiovascular, and respiratory systems (**Table II**). The primary study outcomes were hospital mortality at discharge (n = 9), hospital mortality within 48 hours of admission (n = 3), disposition from the ward or emergency department (pediatric intensive care unit [PICU] admission, n = 2), and the occurrence of a clinical deterioration event, a composite of death, PICU admission, ventilation, cardiopulmonary resuscitation, or use of vasoactive therapy (n = 1).[5,6,12,15-17,30-38]

Risk of Bias Within Studies

The development and validation of all models were at high risk of bias (n = 21 risk of bias assessments for 18 discrete models across 15 studies, **Table III** [available at www.jpeds.com]). The main sources of bias were unclear or poor handling of missing data (eg, complete case analysis, not described, or all missing data treated as normal or abnormal), conversion of continuous variables to dichotomous variables without a priori rationale, issues with modelling methods (eg, complete model selection, univariable analysis for selection of candidate variables), insufficient reporting of performance measures (eg, calibration plots or tables), and absence of accounting for model optimism or model complexity (eg, censoring). Although most model evaluations had low concern for applicability (n = 13), 8 had high concern owing to restrictive inclusion criteria, broad exclusion criteria, or inclusion of uncommon laboratory investigations.[15,31,33,34,37]

Prognostic Model Performance

Discrimination. All studies reported discrimination using the C-statistic (n = 15) and the majority reported its 95% CI (n = 10). Discrimination for the outcome of mortality ranged from good to excellent (C-statistic 0.64-0.93) in individual cohorts (**Table IV**).[5,6,12,15-17,30-33,37,38] Two models had excellent discrimination for admission to PICU, the Boston Children's Hospital PEWS and Pediatric Acute Warning Score.[34,35] One model, Pediatric Early Warning Score for Resource-Limited Settings, had excellent discrimination to predict clinical deterioration events.[36]

Calibration. Most studies did not report calibration (n = 11). Four studies reported calibration using the Hosmer-Lemeshow goodness of fit test for 8 models.[15,17,32,33] Poor calibration was reported in one cohort for each the Signs of Inflammation that Kill (SICK), Pediatric Risk of Mortality (PRISM), and early Pediatric Early Death Index in Africa (PEDIA) scores.[15] Only 2 validated models included a calibration table; 1 additional table was available after author contact, and none included a calibration plot.[15]

Classification. Classification measures were reported or could be calculated from 12 studies for 14 prognostic models.[5,6,12,16,17,30-32,34-36,38] Only the early PEDIA score had classification measures reported in more than one cohort for the same score threshold; these were consistent with modest changes in pretest probability (**Table IV**).

Meta-Analysis

Seven prognostic models were externally validated in >1 cohort: early, immediate, and late PEDIA, Lambarene Organ Dysfunction Score (LODS), Bedside Pediatric Early Warning Score (Bedside PEWS), SICK, and PRISM.[5,6,12,15,17,32] Variables to calculate the complete PRISM score were not available in 2 cohorts; therefore, this model was not included

Table I. Prognostic model development and validation study characteristics

Authors, country	Publication year	Study design	Population	Prediction model	Outcome	Sample size	Outcome events	Events per candidate variable
Development studies								
Bains et al,[38] India	2012	Prospective cohort	Children attending an emergency department	TOPRS	Hospital mortality	777	127	22
Berkley et al,[12] Kenya	2003	Prospective cohort	Children >90 days admitted to hospital Exclusions: trauma and elective procedures	PEDIA	Hospital mortality at 4, 48 and >48 h	8091	436	–
George et al,[20] Tanzania, Uganda, Kenya	2015	Retrospective cohort from randomized controlled trial	Children 2 mo to 12 y with abnormal temperature and impaired perfusion Exclusions: trauma, gastroenteritis, severe malnutrition, burns, non-infectious illness	FEAST PET PETaL	Hospital mortality at 48 h	1044	72	12.6
Kumar et al,[30] India	2003	Prospective cohort	Consecutive pediatric patients admitted to the ward or intensive care unit Exclusions: left against medical advice, transfer to another hospital	SICK	Hospital mortality	1099	44	5.5
Lowlaavar et al,[31] Uganda	2016	Prospective cohort	Admitted children 6-60 mo with proven or suspected infection	Unnamed	Hospital mortality	1307	65	2.7
Mpimbaza et al,[6] Uganda	2015	Prospective cohort	Children admitted to 4 public hospitals in districts of Uganda	Unnamed	Hospital mortality	50 249	1742	72.5
Olson et al,[16] Malawi	2013	Nested case-control	Children aged <15 y admitted to acute care and malnutrition wards during rainy season Exclusions: children on surgical and high dependency units	ITAT	48-h mortality	161	54	13.5

(*continued*)

Table I. Continued

Authors, country	Publication year	Study design	Population	Prediction model	Outcome	Sample size	Outcome events	Events per candidate variable
Rosman et al,[36] Rwanda	2019	Nested case-control	Children from 0-18 y admitted to general pediatric ward for >24 h Exclusions: Deterioration in first 24 h, NICU patients	PEWS-RL	Clinical deterioration event	138	68	–
Validation studies								
Agulnik et al,[34] Guatemala	2017	Retrospective case-control study	Children admitted to a general ward or PICU in a pediatric oncology hospital	BCH PEWS	Unplanned admission to PICU	258	129	–
Chaiyakulsil et al,[35] Thailand	2015	Prospective cohort	Children presenting to the ED Exclusions: trauma, psychiatry, dental, surgical patients	PEWS (PAWS)	ED disposition (ICU)	1136	6	–
Berkley et al,[12] Kenya	2003	Prospective cohort	Children >90 days admitted to hospital Exclusions: trauma and elective procedures	Immediate, Early, Late PEDIA	Hospital mortality at 4 h, 48 h, and >48 h	4802	222	–
George et al,[20] Tanzania, Uganda, Kenya	2015	Retrospective cohort (control arm of randomized controlled trial)	Children 2 mo to 12 y with abnormal temperature and impaired perfusion. Exclusions: trauma, gastroenteritis, severe malnutrition, burns, non-infectious illness	Bedside PEWS PEDIA, PRISM, LODS, FEAST PET(aL)	48-h hospital mortality	1: 1044 2: 5098	1: 72 2: 117	–
Conroy et al,[17] Uganda	2015	Prospective cohort	Admitted children 2 mo to 5 y with fever in prior 48 h	SICK, LODS, early PEDIA	48-h and 7-d hospital mortality	2089	99	–
Gupta et al,[32] India	2010	Prospective cohort	Children 1 mo to 12 y referred to the pediatric team from the ED Exclusions: left against medical advice and referral to another hospital	SICK	Hospital mortality	3895	58	–
Gérardin et al,[33] Senegal	2006	Prospective cohort	Children 0-15 y hospitalized with clinical malaria	PRISM	Hospital mortality	311	28	–

(*continued*)

Table I. Continued

Authors, country	Publication year	Study design	Population	Prediction model	Outcome	Sample size	Outcome events	Events per candidate variable
Mpimbaza et al,[6] Uganda	2015	Prospective cohort	Children admitted to 4 public hospitals in districts of Uganda	Unnamed	Hospital mortality	20 406	556	–
Nariadhara et al,[5] Tanzania	2019	Prospective cohort	Children aged 28 d to 5 y presenting to a tertiary emergency department	mSIRS	24-h and hospital mortality	1350	107	–
Xie et al,[37] China	2019	Nested case control	Children 28 d to 9 y admitted to hospital with a diagnosis of sepsis Exclusions: oncology and immunology patients	Brighton PEWS	Hospital mortality	96	48	–

–, not reported; *BCH*, Boston Children Hospital; *ED*, emergency department; *Feast PET*, Feast Pediatric Emergency Triage Score; *ICU*, intensive care unit; *ITAT*, Inpatient Triage and Treatment; *mSIRS*, Modified Systemic Inflammatory Syndrome; *NICU*, neonatal intensive care unit; *PAWS*, Pediatric Acute Warning Score; *PEWS-RL*, Pediatric Early Warning Score for Resource-Limited Settings.

in the meta-analysis. None of the prognostic models for clinical deterioration events was validated in >1 cohort (**Table IV**). The LODS score had the highest summary C-statistic, 0.85 (95% CI, 0.77-0.92), with high heterogeneity (I^2 = 94%). The SICK score performed similarly with a summary C-statistic of 0.85 (95% CI, 0.82-0.88) with no heterogeneity (I^2 = 0%).

Bias Across Studies

All models were initially rated as moderate quality but downgraded owing to a high risk of bias (primarily related to modelling methods and incomplete reporting). Other models were further downgraded owing to lack of relevance to the complete review question (eg, population inclusion criteria) and inconsistency of measured discriminative ability.[15,17] The overall quality ratings were low to very low (**Table V**; available at www.jpeds.com).

Discussion

This systematic review considered all pediatric prognostic models for risk of mortality and clinical deterioration events developed or validated in LMICs from 2000 to 2019. We included 15 studies describing 15 discrete prognostic models for hospital mortality and 3 for clinical deterioration events; 13 of these studies were externally validated, of which 7 were validated in >1 external cohort: SICK, LODS, early PEDIA, immediate PEDIA, late PEDIA, Bedside PEWS, and PRISM. Discrimination of individual mortality prognostic models ranged from good to excellent (0.64-0.93). Calibration was incompletely reported.

The findings of this systematic review will be of interest to clinicians or researchers working in child health in LMICs and highlight the difficulties in reliably identifying children at risk of poor outcomes. Strengths include a broad search of databases relevant to the global child health literature, a detailed examination of prognostic model development and validation methodology, and use of the recently published PROBAST risk of bias tool. The review is limited by low event rates in some studies, a high degree of bias, and a lack of available data. In several studies, there was incomplete reporting of the rationale for candidate predictor selection, method for variable selection in the final model, handling of missing data, censoring of patients lost to follow-up, and adjusting for model optimism. Only 3 studies reported the use of multiple imputation for handling of missing data, as is recommended by PROBAST.[15,16,31] Although an event per variable rate of >10 for development studies and >100 events for validation studies is recommended, 3 of 7 development cohorts and 4 of 9 validation cohorts did not reach this goal.[21,30-33,35,37] Performance characteristics were often omitted, including calibration and classification measures. The overall quality of the evidence was low to very low.

There were several challenges to integrating the included studies in meta-analyses. First, significant overlap was noted among predictor variables retained in the final models,

Table II. Final model predictor variables

Variables	ITAT	LODS	Lowlaavar	Mpimbaza	mSIRS	PEDIA*	FEAST PET(aL)	PRISM†	SICK	TOPRS	Bedside PEWS	BCH PEWS	Brighton Pews	PAWS	PEWS-RL
Predictor variable															
Nutritional status															
Weight-for-age Z score			X												
Kwashiorkor						X									
Wasting						X									
Neurologic															
Behavior													X		
Blantyre Coma Scale			X												
Impaired consciousness				X		X	X	X							
"Loss of sensorium"										X					
Abnormal mental status															X
Convulsions				X		X				X					
AVPU score									X			X		X	
Coma		X		X											
Prostration		X		X		X									
Unable to sit/stand				X											
Signs of meningitis				X											
Pupillary reflexes								X							
Cardiovascular signs															
Skin color				X			X					X	X		
CRT							X		X		X	X	X		
Systolic blood pressure					X			X	X		X			X	
Peripheral pulses							X					X			
Heart rate	X				X		X	X	X	X	X		X	X	X
Temperature	X			X	X	X	X	X	X	X				X	X
Respiratory signs															
Respiratory rate	X				X				X	X	X	X	X	X	X
Work of breathing				X		X	X				X	X	X	X	X
Supplemental oxygen											X	X	X		X
Oxygen saturation	X		X						X	X	X	X		X	
Deep breathing		X		X		X									
Lung crepitations							X								
Other															
Age				X					X						
Jaundice				X		X									
HIV diagnosis			X												
Parental concern												X			
Nursing concern												X			
History >7 days						X									
Laboratory findings															
Hemoglobin <50 g/L						X									
White blood cell count					X										
Blood glucose								X							
Lactate							(X)	X							
pH <7.2							(X)	X							
BUN >20 mg/dL							(X)								
pCO_2								X							

AVPU, alert, voice, pain, unresponsive; *pCO_2*, partial pressure of carbon dioxide;
*The immediate, early, and late PEDIA use different combination of these indicators.
†Additional PRISM score variables used by Gerardin et al but not George et al: pupillary reflexes, pH, total CO_2, PCO_2, arterial PaO_2, creatinine, urea, white blood cell count, partial thromboplastin time, and platelets.

Table IV. Performance characteristics of prognostic models

Model names and authors	Year	Outcome	C-statistic (95% CI)	Hosmer-Lemeshow *P* value	Cut-off point	Sensitivity (95% CI)	Specificity (95% CI)	LR + (95% CI)
Mortality prognostic models								
Bedside PEWS								
George et al[15] cohort 1*,†	2015	48-h mortality	0.64 (0.56-0.71)	.46	–	–	–	–
George et al[15] cohort 2*,†	2015	48-h mortality	0.74 (0.69-0.79)	.22	–	–	–	–
Summary statistic			**0.69 (0.6-0.79); I^2 = 79%**	–	–	–	–	–
Brighton PEWS								
Xie et al[37]	2019	Hospital mortality	0.77 (–)	–	6.5	74.5 (–)	68.1 (–)	-
FEAST PET								
George et al[15] cohort 1	2015	48-h mortality	0.82 (0.77-0.87)	.56	–	–	–	–
George et al[15] cohort 2	2015	48-h mortality	0.86 (0.82-0.89)	.5	–	–	–	–
FEAST PETaL								
George et al[15] cohort 1	2015	48-h mortality	0.86 (0.82-0.90)	–	–	–	–	–
ITAT			–	–	–	–	–	–
Olson et al[16]	2013	48-h mortality	0.76 (–)	–	≥4	0.44 (–)	0.86 (–)	1.7 (–)
LODS	2015							
Conroy et al[17]	2015	48-h mortality	0.90 (0.88-0.91)	.789	>1	0.82 (0.73-0.89)	0.87 (0.86-0.89)	6.5 (5.6-7.6)
George et al[15] cohort 1	2015	48-h mortality	0.77 (0.72-0.82)	.38	–	–	–	–
George et al[15] cohort 2	2015	48-h mortality	0.87 (0.83-0.90)	.74	–	–	–	–
Summary statistic			**0.85 (0.77-0.92); I^2 = 94%**	–	–	–	–	–
MSIRS								
Nariadhara et al[5]	2019	Hospital mortality	0.63 (0.57-0.68)	–	≥2	0.66 (–)	0.62 (–)	1.78 (–)
Nariadhara et al[5]	2019	24-h mortality	0.70 (0.61-0.79)	–	≥2	0.82 (–)	0.61 (–)	2.12 (–)
Early PEDIA								
Berkley et al[12]	2003	48-h mortality	0.82 (0.80-0.83)	–	>2	0.69 (0.59-0.79)	0.82 (0.81-0.83)	3.8(3.7-4.4)
Conroy et al[17]	2015	48-h mortality	0.90 (0.88-0.91)	.22	>2	0.87 (0.79-0.97)	0.80 (0.79-0.82)	4.4 (3.9-5.0)
George et al[15] cohort 1	2015	48-h hospital mortality	0.70 (0.63-0.77)	.02	–	–	–	–
George et al[15] cohort 2	2015	2-day hospital mortality	0.84 (0.78-0.89)	.08	–	–	–	–
Summary statistic			**0.82 (0.74-0.89); I^2 = 98%**		**>2**	**0.78 (0.61-0.96); $I^2$85%**	**0.81 (0.79-0.83); I^2 = 78.9%**	**4.1 (3.5-4.7); I^2 = 53.8%**
Immediate PEDIA								
Berkley et al[12]	2003	4-h mortality	0.93 (0.92-0.94)	–	–	–	–	–
George et al[15] cohort 1	2015	First-day mortality	0.75 (0.68-0.83)	.64	–	–	–	–
George et al[15] cohort 2	2015	First-day mortality	0.89 (0.84-0.94)	.15	–	–	–	–
Summary statistic			**0.86 (0.76-0.97); I^2 = 94%**		–	–	–	–
Late PEDIA								
Berkley et al[12]	2003	Hospital mortality	0.82 (0.81-0.84)	–	–	–	–	–
George et al[15] cohort 1	2015	Hospital mortality after 2 d	0.55 (0.40-0.69)	.35	–	–	–	–

(*continued*)

Table IV. Continued

Model names and authors	Year	Outcome	C-statistic (95% CI)	Hosmer-Lemeshow *P* value	Cut-off point	Sensitivity (95% CI)	Specificity (95% CI)	LR + (95% CI)
George et al[15] cohort 2	2015	Hospital mortality after 2 d	0.72 (0.66-0.77)	.08	–	–	–	–
Summary statistic			**0.71 (0.56-0.86); I^2 = 95%**					
PRISM‡								
Gérardin et al[33]	2006	Hospital mortality	0.89 (0.85-0.92)	<.001	–	–	–	–
George et al[15] cohort 1	2015	48-h mortality	0.71 (0.61-0.81)	.26	–	–	–	–
George et al[15] cohort 2	2015	48-h mortality	0.77 (0.73-0.82)	.01				
SICK								
Conroy et al[17]	2015	48-h mortality	0.85 (0.83-0.86)	.007	≥2.5	0.82 (0.73-0.89)	0.82 (0.80-0.83)	4.4 (3.9-5.0)
Gupta et al[32]	2010	Hospital mortality	0.85 (0.78-0.92)	.345	≥2	0.76 (0.63-0.86)	0.89(0.88-0.90)	6.7(5.7-8.0)
Kumar et al[30]		Hospital mortality	0.89 (–)	–	≥2.5	0.8(–)	0.82 (–)	4.7 (–)
Summary statistic§			0.85 (0.82 - 0.88); I^2 = 0%					
TOPRS								
Bains et al[38]	2012	Hospital mortality	0.82 (–)	–	≥2.5	0.80(–)	0.74 (–)	3.1(–)
Unnamed score 1	2016							
Lowlaavar et al[31]	2016	Hospital mortality	0.85 (0.80-0.89)	–		0.83 (0.74-0.92)	0.76 (0.73-0.78)	3.5 (–)
Unnamed score 2¶				–				
Mpimbaza et al[6] Derivation cohort	2015	Hospital mortality	0.76 (–)	–	High risk	0.28 (0.26-0.30)	0.95 (0.95-0.92)	5.6 (5.1-6.1)
Mpimbaza et al[6] Validation cohort	2015	Hospital mortality	0.74 (–)	–	High risk	0.18 (0.15-0.22)	0.96 (0.96-0.97)	5.1 (4.2-6.2)
Prognostic models for clinical deterioration events								
BCH PEWS								
Agulnik et al[34]	2017	Admission to PICU	0.94 (0.91-0.97)	–	≥4	0.88 (–)	0.93 (–)	12.8 (–)
PAWS								
Chaiyakulsil et al[35]	2015	Admission to PICU	0.98 (0.96-1.0)	–	≥3	1 (–)	0.91 (–)	10.5 (–)
PEWS-RL								
Rosman et al[36]	2019	Clinical deterioration event	0.96 (0.93-0.99)	–	≥3	0.96 (–)	0.87 (–)	7.57 (–)

– Data not included and not possible to calculate from primary study.
*Cohort 1: derivation dataset from the FEAST trial control arm; cohort 2, validation dataset from the Kilifi pediatric ward.
†Systolic blood pressure not available in validation data for Bedside PEWS calculation.
‡No summary statistic provided as complete score not used by George et al.
§Summary classification measures not calculated as different score thresholds were used.
¶No summary statistic as confidence intervals and SE not included or possible to calculate.

although often these were defined differently. For example, 13 scores included a description of level of consciousness using many different terms. Although these terms may represent similar concepts, the differences render comparison across models challenging. Second, a comparison of discrimination of the 6 prognostic models demonstrated the LODS score to have the best discrimination (summary C-statistic,0.85; 95% CI, 0.77-0.92). However, the high degree of heterogeneity (I^2 = 94%) may be related to differences in study inclusion criteria regarding patient age, presence of fever, and signs of impaired perfusion. An insufficient number of studies precluded a metaregression to examine the impact of different inclusion criteria on model performance. In contrast, the SICK score performed similarly well with a summary C-statistic of 0.85 (95% CI, 0.82-0.88) and consistently (I^2 = 0%). As expected, prognostic models had improved discrimination in more heterogeneous populations than homogenous ones; for example, the Fluid Expansion As Supportive Therapy Pediatric Emergency Triage and Laboratory score had a higher C-statistic in the general pediatric validation cohort than in the control arm of a randomized controlled trial with restrictive inclusion criteria.[39]

The PEWS models deserve separate mention given that an elevated PEWS is intended to prompt treatment escalation. Therefore, the outcome of PICU admission may be self-fulfilling in that children with an elevated PEWS are more likely to trigger consultation to the PICU. Two PEWS prognostic models had good discrimination, but calibration was not reported.[15,34,35] The third PEWS score—the PEWS for Resource-Limited Settings—reported excellent discrimination for a composite outcome of clinical deterioration events, of which 87% were deaths.[36] In a cohort of children with sepsis, the Brighton PEWS had moderate discrimination for hospital mortality (area under the curve, 0.77; no CI reported).[37] In another cohort of children admitted to a pediatric ward, the Bedside PEWS had poor to moderate discrimination for hospital mortality, which may be related to missing data.[15]

In light of these limitations, further study is required before any pediatric prognostic mortality or clinical deterioration event model can be recommended for research or clinical practice. Additional research should first focus on rigorous validation and updating of existing models with good discrimination, rather than development of new models.[40] Inclusion criteria should be broad and recalibration may be necessary such that study results are generalizable to pediatric emergency care and general admission contexts. Researchers should follow recent recommendations for the conduct and evaluation of prediction model validation studies to ensure complete reporting of discrimination, calibration and classification measures.[41-43] Next, studies should evaluate both the feasibility of implementing models in high-volume, low-resource clinical settings and the effect of implementation on clinically important outcomes. Many models were easier to implement in LMICs owing to a condensed set of clinical predictor variables. However, even in a research context, up to 10% of predictor data were missing, which would likely be higher in clinical practice.[17,35] In addition, score cut-offs should be specified to classify children at high or low risk of death, because both categories have implications for triage and disposition. Clinicians would be most aided by a tool with a limited number of easily measured clinical signs, highly discriminative for poor outcomes, and responsive to changes in clinical status following therapeutic interventions. Finally, qualitative evaluation of the perceived usefulness and barriers to uptake of prediction models would facilitate their development.[44]

Several scores were identified with overlapping features and comparable discrimination ability. Overall quality of evidence was low owing to restrictive inclusion criteria and concerns regarding analysis methods. Few studies were validated in >1 external cohort; thus, there is a clear need for further validation studies before impact analysis or widescale implementation. Further research would benefit from improved handling of missing data, complete reporting of calibration and classification measures and sufficient sample size. ■

Submitted for publication Jan 22, 2020; last revision received Apr 11, 2020; accepted May 12, 2020.

Reprint requests: Fiona Muttalib, MD, Center for Global Child Health, Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada. E-mail: Fiona.muttalib@mail.utoronto.ca

References

1. United Nations Inter-agency Group for Child Mortality Estimation. Levels and trends in child mortality: report 2018. Estimates developed by the United Nations Inter-agency Group for Child Mortality Estimation. New York: 2018.
2. English M. Child survival: district hospitals and paediatricians. Arch Dis Child 2005;90:974-8.
3. Campbell H, Duke T, Weber M, English M, Carai S, Tamburlini G, et al. Global initiatives for improving hospital care for children: state of the art and future prospects. Pediatrics 2008;121:e984-92.
4. Li MY, Kelly J, Subhi R, Were W, Duke T. Global use of the WHO pocket book of hospital care for children. Paediatr Int Child Health 2013;33:4-11.
5. Nariadhara MR, Sawe HR, Runyon MS, Mwafongo V, Murray BL. Modified systemic inflammatory response syndrome and provider gestalt predicting adverse outcomes in children under 5 years presenting to an urban emergency department of a tertiary hospital in Tanzania. Trop Med Health 2019;47:13.
6. Mpimbaza A, Sears D, Sserwanga A, Kigozi R, Rubahika D, Nadler A, et al. Admission risk score to predict inpatient pediatric mortality at four public hospitals in Uganda; 2015. p. 0133950.
7. English M, Esamai F, Wasunna A, Were F, Ogutu B, Wamae A, et al. Assessment of inpatient paediatric care in first referral level hospitals in 13 districts in Kenya. Lancet 2004;363:1948-53.
8. 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;319:1002-12.
9. Feudtner C, Berry JG, Parry G, Hain P, Morse RB, Slonim AD, et al. Statistical uncertainty of mortality rates and rankings for children's hospitals. Pediatrics 2011;128:e966-72.
10. Olson D, Preidis GA, Milazi R, Spinler JK, Lufesi N, Mwansambo C, et al. Task shifting an inpatient triage, assessment and treatment programme

improves the quality of care for hospitalised Malawian children. Trop Med Int Health 2013;18:879-86.
11. Molyneux E. Paediatric emergency care in developing countries. Lancet 2001;35786-7.
12. Berkley JA, Ross A, Mwangi I, Osier FH, Mohammed M, Shebbe M, et al. Prognostic indicators of early and late death in children admitted to district hospital in Kenya: cohort study. BMJ 2003;326:361.
13. Ayieko P, Ogero M, Makone B, Julius T, Mbevi G, Nyachiro W, et al. Characteristics of admissions and variations in the use of basic investigations, treatments and outcomes in Kenyan hospitals within a new Clinical Information Network. Arch Dis Child 2016;101:223-9.
14. Wiens MO. Childhood mortality from acute infectious diseases in Uganda: studies in sepsis and post-discharge mortality. Vancouver: University of British Columbia; 2015.
15. George EC, Walker AS, Kiguli S, Olupot-Olupot P, Opoka RO, Engoru C, et al. Predicting mortality in sick African children: the FEAST Paediatric Emergency Triage (PET) Score. BMC Med 2015;13:174.
16. Olson D, Davis NL, Milazi R, Lufesi N, Miller WC, Preidis GA, et al. Development of a severity of illness scoring system (inpatient triage, assessment and treatment) for resource-constrained hospitals in developing countries. Trop Med Int Health 2013;18:871-8.
17. Conroy AL, Hawkes M, Hayford K, Namasopo S, Opoka RO, John CC, et al. Prospective validation of pediatric disease severity scores to predict mortality in Ugandan children presenting with malaria and non-malaria febrile illness. Crit Care 2015;19:47.
18. Brown SR, Martinez Garcia D, Agulnik A. Scoping review of Pediatric Early Warning Systems (PEWS) in resource-limited and humanitarian settings. Front Pediatr 2018;6:410.
19. World Bank Group. World Bank Country and Lending Groups 2018. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups. Accessed October 28, 2019.
20. Agulnik A, Mora Robles LN, Forbes PW, Soberanis Vasquez DJ, Mack R, Antillon-Klussmann F, et al. Improved outcomes after successful implementation of a pediatric early warning system (PEWS) in a resource-limited pediatric oncology hospital. Cancer 2017;123:2965-74.
21. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11:e1001744.
22. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51-8.
23. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res 2018;28:2768-86.
24. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA 2017;318:1377-84.
25. Debray TP, de Jong V. Diagnostic and prognostic meta-analysis. R package version 0.2.0; 2019.
26. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
27. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002.
28. Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. Syst Rev 2013;2:71.
29. Iorio A, Spencer F, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ 2015;350:h870.
30. Kumar N, Thomas N, Singhal D, Puliyel JM, Sreenivas V. Triage score for severity of illness. Indian Pediatr 2003;40:204-10.
31. Lowlaavar N, Larson CP, Kumbakumba E, Zhou G, Ansermino JM, Singer J, et al. Pediatric in-hospital death from infectious disease in Uganda: derivation of clinical prediction models. PLoS One 2016;11:e0150683.
32. Gupta MA, Chakrabarty A, Halstead R, Sahni M, Rangasami J, Puliyel A, et al. Validation of "Signs of Inflammation in Children that Kill" (SICK) score for immediate non-invasive assessment of severity of illness. Ital J Pediatr 2010;36:35.
33. Gerardin P, Rogier C, Leteurtre S, Jouvencel P, Ka AS, Imbert P. Evaluation of Pediatric Risk of Mortality (PRISM) scoring in African children with falciparum malaria. Pediatr Crit Care Med 2006;7:45-7.
34. Agulnik A, Mendez Aceituno A, Mora Robles LN, Forbes PW, Soberanis Vasquez DJ, Mack R, et al. Validation of a pediatric early warning system for hospitalized pediatric oncology patients in a resource-limited setting. Cancer 2017;123:4903-13.
35. Chaiyakulsil C, Pandee U. Validation of pediatric early warning score in pediatric emergency department. Pediatr Int 2015;57:694-8.
36. Rosman SL, Karangwa V, Law M, Monuteaux MC, Briscoe CD, McCall N. Provisional validation of a pediatric early warning score for resource-limited settings. Pediatrics 2019;143:e20183657.
37. Xie X, Li M, Xiong TT, Wang R, Xiao L. Nested case-control study of multiple serological indexes and Brighton pediatric early warming score in predicting death of children with sepsis. World J Clin Cases 2019;7:431-40.
38. Bains HS, Soni RK. A simple clinical score "TOPRS" to predict outcome in pediatric emergency department in a teaching hospital in India. Iran J Pediatr 2012;22:97-101.
39. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017;356:i6460.
40. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012;98:691-8.
41. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: explanation and elaboration. Ann Intern Med 2019;170:W1-33.
42. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg 2015;102:148-58.
43. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 2014;35:1925-31.
44. Kappen TH, van Loon K, Kappen MA, van Wolfswinkel L, Vergouwe Y, van Klei WA, et al. Barriers and facilitators perceived by physicians when using prediction models in practice. J Clin Epidemiol 2016;70:136-45.

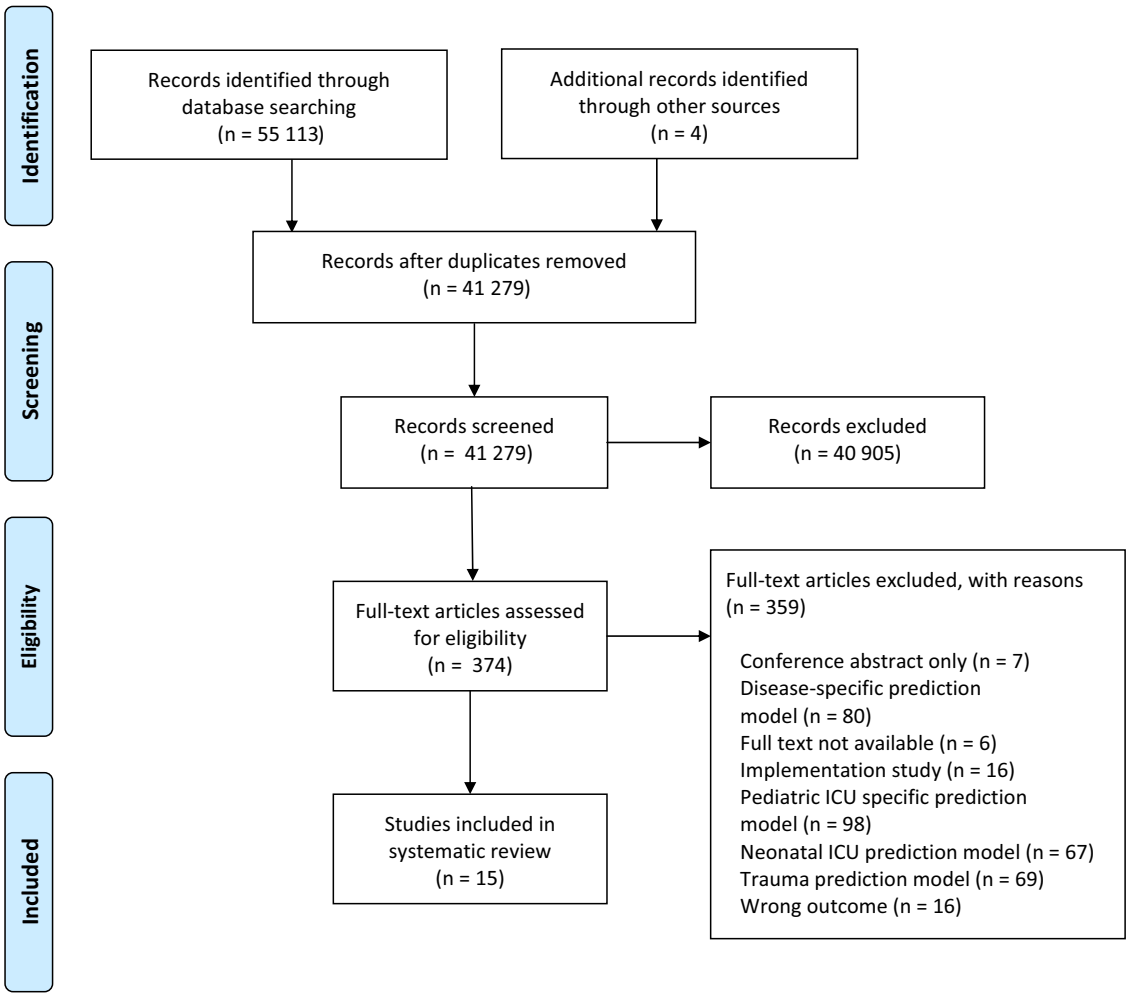

Figure. PRISMA flow diagram of included studies. *ICU*, intensive care unit.

Table III. Risk of bias assessment using the PROBAST tool

Study and models	Year	Risk of bias					Applicability			
		Population	Predictors	Outcome	Analysis	Overall	Population	Predictors	Outcome	Overall
Agulnik, BCH PEWS	2017	+	?	+	+	+	+	–	–	+
Bains, TOPRS	2012	+	–	–	+	+	–	-	–	–
Berkley, PEDIA	2003	–	–	–	+	+	–	–	–	–
Chaiyakulsil, PAWS	2015	–	–	–	+	+	–	–	+	+
Conroy, PEDIA	2015	–	–	–	+	+	–	–	–	–
Conroy, SICK	2015	–	–	–	+	+	–	–	–	–
Conroy, LODS	2015	–	–	–	+	+	–	–	–	–
George, FEAST PET	2015	+	–	–	+	+	+	–	–	+
George, PEDIA	2015	+	+	–	+	+	+	–	–	+
George, PRISM	2015	+	+	–	+	+	+	+	–	+
George, LODS	2015	+	–	–	+	+	+	–	–	+
George, Bedside PEWS	2015	+	–	–	+	+	+	–	–	+
Gerardin PRISM	2006	?	?	?	+	+	+	?	?	+
Gupta, SICK	2010	–	–	?	+	+	–	–	–	–
Kumar, SICK	2003	–	–	–	+	+	–	–	–	–
Lowlaavar, Unnamed	2016	–	–	–	+	+	+	–	–	+
Mpimbaza, Unnamed	2015	–	–	–	+	+	–	–	–	–
Nariadarah, MSIRS	2019	–	–	–	+	+	–	–	–	–
Olson, ITAT	2013	–	–	–	+	+	–	–	–	–
Rosman, PEWS-RL	2019	+	–	?	+	+	+	–	–	+
Xie, Brighton PEWS	2019	+	–	–	+	+	+	–	–	+

ITAT, Inpatient triage and treatment; *mSIRS*, Modified Systemic Inflammatory Syndrome; *PEWS-RL*, Pediatric Early Warning Score for Resource-Limited Settings.
– indicates low risk of bias/low concern regarding applicability; + indicates high risk of bias/high concern regarding applicability; and ? indicates unclear risk of bias/unclear concern regarding applicability.

Table V. Risk of bias across studies

Prognostic models	No. of studies	No. of cohorts	Estimated effect size: discrimination (95% CI)	Phase	Study limitations	Inconsistency	Indirectness	Imprecision	Moderate/large effect size (discrimination calibration)	Dose effect	Overall quality
Mortality											
Bedside PEWS	1	2	0.69 (0.60-0.79)	2	++	+	+	–	?	?	Very low
FEAST PET	1	2	0.84 (0.81-0.88)	2	++	–	+	–	?	?	Very Low
LODS	2	3	0.85 (0.77-0.92)	2	++	+	–	–	?	?	Very Low
Mpimbaza	1	2	0.75 (?)	2	++	–	–	?	?	–	Low
Early PEDIA	3	4	0.82 (0.74-0.89)	2	++	+	–	–	?	–	Low
Late PEDIA	2	2	0.82 (0.80-0.82)	2	++	–	–	–	?	–	Low
SICK	3	3	0.85 (0.82-0.88)	2	++	–	–	+	?	–	Low

?, unknown; +, serious limitations or presence of inconsistency, indirectness, imprecision; ++, very serious limitations; –, absence of inconsistency, indirectness, imprecision or presence of a dose effect and moderate/large effect size;