Averroes European Medical Journal
Open Access Publisher

November 2026 • Volume 1 • Number 1-ENGV

Development, Validation, and Clinical Applicability of Predictive Models for Pediatric Sepsis: A Critical Systematic Review of the Literature (2021–2026)

Abstract

Background. Pediatric sepsis remains a leading cause of preventable death. The 2024 international consensus revision removed systemic inflammatory response syndrome from the diagnostic framework and adopted a definition based on organ dysfunction associated with suspected or confirmed infection, operationalised by the Phoenix Sepsis Score. This recalibration shifts the clinical problem from the recognition of established organ dysfunction to its anticipation [3,4,5,6].

Objective. To appraise systematically the studies addressing the development and validation of predictive models for sepsis in children and neonates, characterising their discriminative performance, the robustness of validation, and the clinical transferability of the proposed instruments. Sources and selection.

PubMed/MEDLINE was searched (January 2021 – November 2026).

Derivation studies, external validations, evaluations of devices in clinical deployment, and systematic reviews, meta-analyses and scoping reviews were eligible. Thirty-four studies were included. Data were extracted using a structured framework: source population and level of care; operational definition of sepsis; type of model employed and number of covariates; measures of discrimination and calibration with their confidence intervals; level of validation; and identified sources of bias.

Synthesis of findings. An umbrella review of 18 systematic reviews reports a pooled sensitivity of 0.81 (95% CI 0.76–0.85) and a specificity of 0.79 (95% CI 0.74–0.84), with areas under the curve ranging from 0.78 to 0.94 according to resource setting [1].

Individual studies range from 0.56 to 0.99 [22].

Three quarters of neonatal models (34 of 44) have undergone only a single validation [23]. Evaluation of a clinically deployed device, encompassing 166,943 emergency department encounters, demonstrated a sensitivity of 36% (95% CI 29–43%) and a positive predictive value of 3%, with lower sensitivity among children of sub-Saharan African ancestry, children under 5 years of age, and those managed outside tertiary pediatric centres [2].

Conclusion. The dissociation between performance measured in research settings and that observed in deployment — from 94% to 36% sensitivity — is the structuring finding of this review. The narrow margin separating complex algorithms from parsimonious models indicates that predictive performance depends less on algorithmic complexity than on the quality of covariates, the repetition of external validations, and equity of performance across subpopulations.

Keywords: pediatric sepsis; septic shock; early detection; supervised learning; Phoenix Sepsis Score; external validation; health equity; endothelial biomarkers.

Author contributions

Ahmed HADDADI[1]; Karima HADDADI[2]; Souad HADDADI[3]

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

1. Introduction

Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection. In pediatric practice, it remains one of the leading causes of death worldwide, with the burden concentrated during the neonatal period and in low- and middle-income countries [3,4].

The structural difficulty characterising sepsis is diagnostic. In children, sepsis initially manifests with non-specific features — fever, tachycardia, tachypnoea, altered mental status — that are indistinguishable from those of uncomplicated infection, yet progression to septic shock may occur within hours.

This characteristic makes delayed diagnosis the principal determinant of preventable mortality. Systemic inflammatory response syndrome served as a diagnostic criterion for three decades and was formally discarded in 2024: its inadequate sensitivity and specificity disqualify it for establishing a diagnosis of sepsis in children [5]. International consensus now rests on the concept of organ dysfunction associated with suspected or confirmed infection, operationalised by a Phoenix Sepsis Score of at least 2 points. This composite score encompasses four domains — cardiovascular, respiratory, neurological and haemostatic [6]. It was developed and validated in more than 3.6 million pediatric inpatient encounters across 10 sites in 5 countries [6]. The Phoenix Sepsis Score, however, identifies organ dysfunction that is already established. The clinical challenge lies in anticipating it. It is this shift — from recognition to prediction — that structures the literature of the past five years and governs the interpretation of the findings presented below.

2. Materials and methods

2.1. Literature search

PubMed/MEDLINE was searched for the period January 2021 – November 2026. Search strategies combined, in the title–abstract and full-text fields, terms relating to pediatric and neonatal sepsis (pediatric sepsis, neonatal sepsis, septic shock) with those relating to prediction and modelling (prediction model, clinical prediction rule, machine learning, early warning score, external validation). A complementary manual search of the reference lists of the included studies was performed.

2.2. Eligibility criteria

Inclusion criteria. Derivation studies of predictive models; external validation studies, including temporal validations; evaluations of devices in clinical deployment; systematic reviews, meta-analyses and scoping reviews addressing prediction of sepsis in children or neonates.

Exclusion criteria. Studies without assessment of predictive performance; purely descriptive or epidemiological studies; publications without usable measures of discrimination or calibration; models developed and evaluated exclusively in adults.

2.3. Data extraction and critical appraisal

The variables extracted were: source population and level of care; operational definition of sepsis; type of model employed and number of covariates; measures of discrimination and calibration with their confidence intervals; level of validation (internal, external, temporal or deployment); and sources of bias reported by the authors or identified during critical appraisal.

A structured appraisal framework was applied to each study across five dimensions: consistency of the outcome definition with the 2024 consensus criteria; prevalence of the event in the study population;

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

reproducibility of validation; presence of subgroup analyses addressing equity; and demonstration of clinical benefit. Risk of bias was appraised using the domains of the PROBAST tool (Prediction model Risk Of Bias ASsessment Tool), which distinguishes domains relating to participants, the outcome, and the analysis [52].

3. Results

3.1. Epidemiology: incidence, prevalence and microbiological determinants

Epidemiological characterisation is not a descriptive preliminary: it determines the expected positive predictive value of a model, and therefore its effective clinical utility.

Table 1. Epidemiological indicators of pediatric and neonatal sepsis.

Population	Indicator	Value (95% CI)	Source
Hospitalised children, United States (2016–2023)	Incidence of sepsis	1.3% (51,542/3,925,809)	[7]
	In-hospital mortality	10.1%	[7]
	Community-acquired sepsis	72.6%	[7]
	Sepsis with shock	61.6%	[7]
	Sensitivity of the clinical definition	69.9% (58.1–79.8)	[7]
	Specificity of the clinical definition	93.1% (89.6–95.7)	[7]
Children under 5 years, Africa	Pooled prevalence	40.2% (26.8–55.4)	[8]
	Prevalence adjusted for publication bias	27.1% (15.5–43.0)	[8]
	Pooled mortality	16.3% (10.6–24.4)	[8]
Neonates, low-income countries	Pooled prevalence	40.98% (30.50–51.46)	[9]
Extremely preterm infants (22–28 weeks)	Incidence of early-onset sepsis	3.3% (4,354/130,938)	[10]
	Mortality	35.5%	[10]
	Mortality or morbidity	78.2%	[10]
Neonates, Indian special care units	Gram-negative bacteria	70.0% of isolates	[11]
	Multidrug resistance	75–88% of isolates	[11]
	Case fatality of documented sepsis	36.6% (12.1–71.0)	[11]

This table should be interpreted with two considerations in mind. First, the sensitivities and specificities reported for clinical definitions reflect the performance of the definition itself, not that of a predictive model. Second, prevalences derived from resource-limited settings reflect selected hospital populations and are not transferable to emergency departments in high-income countries.

3.2. Discriminative performance: synthesis of studies published 2021–2026

Table 2 summarises the included studies. It represents a juxtaposition of heterogeneous evidence rather than a direct comparison: populations, operational definitions of the outcome, and validation procedures differ between studies, precluding any quantitative comparison of performance.

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

Table 2. Included studies: population, type of model employed, performance and level of validation.

Study	Population	Algorithm	Performance	Validation	Equity
Umbrella review, 2024 [1]	18 systematic reviews	Multiple	Sens. 0.81; Spec. 0.79; AUC 0.78–0.94	By resource setting	No
Multicentre study, 2026 [13]	784 children, 2 centres	Random forest, 12 covariates	AUC 0.887 (0.850–0.924); accuracy 82.1%	External	No
Haematological model, 2026 [14]	Neonates and children	CatBoost, 18 covariates	ΔAUC +0.032 to +0.062 vs blood count	2 external cohorts	No
LASSO regression, 2025 [15]	834 children	7 covariates	Adequate discrimination and calibration	Internal	No
Biosensors, 2025 [16]	96 children, Dhaka	LASSO on physiological signals	AUC 0.78 (sepsis); 0.85 (shock); 0.87 (mortality)	Cross-validation	No
Neonatal model, 2026 [17]	Neonates, low- and middle-income countries	Supervised learning	Sens. 90.3%; Spec. 40.6%	Internal	No
Population cohort, 2026 [18]	140,633 children, Hong Kong	Phoenix-8	AUC 0.91 (0.89–0.93)	Population-based	Yes
Saudi validation, 2026 [19]	431 admissions	Phoenix	Sens. 80.9%; PPV 27.0%	External	No
Infant model, 2025 [20]	Children, intensive care	XGBoost	Sens. 0.77; Spec. 0.74	Not specified	No
Index study, 2025 [21]	1,604,422 encounters	Gradient boosting	AUROC 0.94 (0.93–0.94)	Temporal	No
Commercially available device, 2026 [2]	166,943 encounters	Integrated device	Sens. 36%; PPV 3%	Deployment	Yes — asymmetry

First dimension — dispersion of performance and the nature of its determinants. Areas under the curve range from 0.56 to 0.99 [22]. A scoping review of 27 studies reports that 85% were single-centre, 59% used logistic regression, and 74% used datasets with low event prevalence [22]. This dispersion therefore arises predominantly from methodological differences — sample size, prevalence, validation approach — rather than from differences in predictive efficacy. The consequence is direct: a quantitative meta-analysis conducted on these data would measure heterogeneity of methods rather than the intrinsic performance of the algorithms.

Second dimension — magnitude of the gains attributable to algorithmic complexity. Direct comparison between a model based on 18 haematological covariates and a model based on the complete blood count alone yields differences in area under the curve of 0.032 to 0.062, with reversal in one of the external cohorts [14]. A scoping review further reports a mean area of 0.84 for logistic models converted into scores, versus 0.81 for linear models and 0.85 for tree-based models [22]. These differences, spanning three to six hundredths of a point of area under the curve, are of the same order of magnitude as between-sample variability. No conclusion regarding the superiority of any algorithmic family can be drawn from them; they argue in favour of parsimonious models, which are more readily interpretable by the clinician and more consistent with regulatory requirements for transparency.

Third dimension — fragility of validation. A scoping review of neonatal models in low- and middle-income countries identified 82 studies validating 44 distinct models in 24,252 neonates.

Three quarters of the models (34 of 44) had undergone only a single validation; two thirds required laboratory parameters; only four studies (5%) were conducted in the World Health Organization African Region, and only one included data from a low-income country [23].

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

This configuration restricts the transferability of models precisely where the burden — incidence and prevalence — is heaviest. Fourth dimension — deficit in equity analysis.
Of the eleven studies in the table, only one performed a subgroup analysis [2].
This situation does not constitute a merely incidental methodological shortcoming: it deprives of information the very point at which a model's performance carries the most serious clinical consequences.
The four dimensions converge on a single observation: the dispersion of reported performance measures heterogeneity of methods and settings, not a hierarchy of efficacy among algorithms.

3.3. Dissociation between experimental and operational performance

The multicentre evaluation of a clinically deployed — and, in this instance, commercially available — early detection device, encompassing 166,943 pediatric emergency department encounters, demonstrated a sensitivity of 36% (95% CI 29–43%) and a positive predictive value of 3% (95% CI 2–4%) for the identification of sepsis according to the Phoenix Sepsis Score, the alert preceding diagnosis in only 30% of cases [2]. The negative predictive value exceeded 99%.

The stratified analyses constitute the most concerning finding. Sensitivity was at least ten percentage points lower than overall sensitivity among children of sub-Saharan African ancestry (22%; 95% CI 9–40%), among children under 5 years of age (20%; 95% CI 12–30%), and in emergency departments outside tertiary pediatric centres (25%; 95% CI 14–40%) [2]. The authors conclude that the device "does not support equitable or timely pediatric sepsis screening".

Four conclusions follow, which must in certain respects be stated without qualification.
First conclusion. The dissociation between 94% and 36% sensitivity — from the index study to the deployed device — reveals no methodological flaw attributable to the authors of the index study [21]. It reflects the fact that a model developed on structured, complete and regularly updated data will encounter, in routine practice, incomplete and heterogeneous data. Predictive performance is not an intrinsic property of the algorithm: it characterises the algorithm–information environment pair.
Second conclusion. A positive predictive value of 3% combined with a negative predictive value above 99% defines a triage instrument, not a diagnostic one. Its purpose is to exclude patients in whom sepsis is highly improbable. A triage device assessed on its positive predictive value will always appear disappointing; assessed on its negative predictive value, it may offer genuine clinical utility. This distinction has implications for the design of alerting systems and for staff training.
Third conclusion. The asymmetry in performance according to ethnicity, age and type of facility is a matter of health equity. A model that performs less well for the populations most exposed produces the opposite of the intended effect. A 2026 review reports that false-positive rates vary according to level of care, revealing "critical disparities in electronic health record infrastructure" [24].
Fourth conclusion. These findings do not invalidate the predictive approach; they define its conditions of validity. Evaluation of a model on the data of the institution in which it is to be used, together with subgroup analyses, is a prerequisite for deployment and does not duplicate the evaluation conducted during development.

3.4. Pathophysiological foundations of prediction

Any predictive approach presupposes that a detectable phase of dysregulation precedes organ failure. The literature of the past five years supports this premise through five converging lines of evidence.

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

3.4.1. Microvascular phenotypes identified by unsupervised learning

A study of 239 pediatric patients (median age 2.1 years; interquartile range 0.5–11.1) identified three distinct microvascular phenotypes [25]:

Table 3. Microvascular phenotypes of pediatric sepsis and associated outcomes.

Phenotype	n	Microvascular characteristics	MODS at 72 h	Mortality
Preserved	52	High capillary density, mild glycocalyx impairment	10%	1.9%
Partially compensated	75	Marked inflammatory activation, moderate reduction in capillary density	—	12.1%
Decompensated	112	Minimal capillary density, global microvascular failure	42%	17.2%

The risk of multiple organ dysfunction syndrome at 72 hours was highest in the decompensated phenotype compared with the preserved phenotype (adjusted hazard ratio 2.47; 95% CI 1.04–8.44; p = 0.042), independently of age, site of infection and use of invasive mechanical ventilation. Mortality followed a stepwise gradient (1.9%, 12.1%, 17.2%; p = 0.023). Internal validation analyses reproduced both the phenotypic assignment and the prognostic associations [25].

This study provides empirical validation of the predictive premise. The decisive element lies in its independence: the microcirculation carries information that conventional clinical and laboratory covariates do not capture.

3.4.2. Endothelial glycocalyx biomarkers: quantification of a causal mechanism

The mechanism linking inflammatory activation to organ failure is endothelial and quantifiable. A study of 85 septic children and 19 controls matched for age and sex established that serum concentrations of syndecan-1, heparan sulfate and hyaluronic acid were significantly higher in septic patients, and higher in severe than in mild forms; hyaluronic acid was further increased in the presence of shock. In the septic shock cohort, non-survivors had higher concentrations than survivors. On multivariable analysis, syndecan-1, hyaluronic acid and procalcitonin were independent predictors of severe sepsis. In septic shock, a syndecan-1 level of 89.47 ng/mL or above predicted significantly reduced 28-day survival (p = 0.004) [26].

A meta-analysis of 17 studies including 3,529 patients found that patients with elevated syndecan-1 had an increased risk of death (OR 2.32; 95% CI 1.89–3.10), multiple organ dysfunction (OR 3.3; 95% CI 1.51–7.25) and respiratory failure (OR 7.53; 95% CI 1.86–30.45) [27].

Three properties confer predictive value on these markers: they quantify a causal mechanism rather than a non-specific marker of severity; their kinetics precede clinically documented organ failure; and syndecan-1 is measurable in the circulation, permitting its integration into a multimodal model. Their current limitation lies in the absence of prospective validation within a composite model.

3.4.3. Immunological endotypes and identifiable therapeutic targets

A study integrating in-depth immune phenotyping, plasma proteomics, single-cell transcriptomics and phosphoflow cytometry in 88 critically ill children identified three immunological subgroups, including a high-severity group characterised by IL-6- and IFN-γ-dependent hypercytokinaemia. This group displayed T-lymphocyte activation with suppression of a protective lymphoid gene programme in CD8+

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

subpopulations. Phosphoflow cytometry demonstrated basal hyperactivation of STAT1/STAT3 in these lymphocytes, which no longer responded to stimulation [28].

The JAK/STAT axis is thereby identified as a rational therapeutic target [28]. These findings open a precise perspective: prediction could address not the occurrence of sepsis in general, but a determined pathophysiological trajectory, enabling targeted intervention on the dominant mechanism.

A complementary study identified two phenotypes of pediatric septic shock by latent profile analysis, associated with transcriptomic signatures of microvascular endothelial dysfunction and increased neutrophil turnover; these phenotypes showed no overlap with endotypes established on gene expression and displayed no differential response to corticosteroids [29]. This finding indicates that phenotypic and endotypic classifications do not coincide, and that prediction based on one may not capture the information conveyed by the other.

3.4.4. Kinetics of plasma biomarkers and proteomic panels

A study of dynamic biomarkers established that baseline values and early decline of C-reactive protein, procalcitonin, interleukin-6 and D-dimers reflect organ dysfunction and length of hospital stay. Initial inflammatory profiles differed according to aetiology — C-reactive protein and procalcitonin approximately doubled in bacterial compared with viral infections, with interleukin-6 highest in respiratory and abdominal infections. All markers declined significantly over seven days, with uniform proportional declines irrespective of the pathogen. A two-step algorithm — baseline C-reactive protein of 60 mg/L or above, procalcitonin of 3 ng/mL or above, interleukin-6 of 200 pg/mL or above, or D-dimers of 1.5 mg/L or above, combined with a decline of at least 50% in interleukin-6 or procalcitonin within 72 hours — identified children requiring intensive care or hospitalisation of at least seven days with an area under the curve of 0.91 [30]. The value of this finding lies in the accessibility of the markers, which are available in most laboratories, including in resource-constrained settings, in contrast to the research technologies required by multi-omic approaches.

A neonatal serum proteomic study further identified a protein panel spanning three biological processes: pathogen detection (LBP, CD14) and acute-phase inflammation (CRP, SAA1/2, ORM1/2); innate immune activation and leucocyte recruitment (FCGR3A, CSF1R, CD163, CD206) with platelet exhaustion (PF4, PPBP, THBS1, GP5); and endothelial injury with microvascular dysfunction (ICAM1, VCAM1, VWF, SPARC), accompanied by simultaneous loss of protective lipoproteins and serpins [31].

3.4.5. Immune immaturity and the kinetics of organ failure

One determinant conditions the very feasibility of prediction in pediatrics and warrants explicit statement. Compared with adults, neonates display attenuated responses mediated by pattern-recognition receptors (Toll-like receptors) and reduced production of pro-inflammatory cytokines; surface expression of TLR4 on neonatal neutrophils is low, as is that of downstream signal transducers, limiting cytokine responses and chemotaxis [32].

This immaturity does not, however, translate into hyporesponsiveness. Pediatric macrophages produce both more pro-inflammatory TNF-α and more anti-inflammatory IL-10 than adult macrophages; in an ex vivo model of child macrophages stimulated with lipopolysaccharide, the IL-10/TNF-α ratio was fifteen times higher than that observed in adults [32,33]. The clinical translation is direct: children may develop concomitant hyperinflammation and immunosuppression during the first 72 hours of hospitalisation, with a prevalence of simultaneous immune dysfunction higher than that reported in adult cohorts. This

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

overlapping trajectory, together with simultaneous rather than sequential onset of multiple organ dysfunction syndrome, accounts for early mortality significantly higher than in adults [34]. Implication for predictive modelling. The earlier onset of organ failure, simultaneous across several systems, narrows the window for anticipation and blunts the gradual progression of the alert signal. This biological constraint also explains why a model validated in adults cannot be transposed to children without dedicated validation.

3.5. Critical appraisal of methodological limitations

3.5.1. Heterogeneity of operational definitions

The scoping review of predictive technologies notes that "variation in outcome definitions highlights the potential importance of the 2024 consensus criteria in future development" [22]. A model developed and tested on earlier criteria measures the occurrence of a clinical entity that is no longer recognised. Alignment with the Phoenix criteria is therefore not one methodological choice among others: it is a condition of validity. The temporal heterogeneity introduced by the 2024 revision precludes direct comparison of the performance of earlier and later models.

3.5.2. Low prevalence and clinical consequences

Of the 27 studies in the scoping review, 74% used datasets with low prevalence [22]. The consequence is mechanistic: a model with high sensitivity and specificity, applied to a population in which the event is rare, generates a substantial proportion of false alerts — a point confirmed by the 3% positive predictive value measured in deployment [2]. This effect carries a frequently overlooked secondary consequence: the multiplication of false alerts induces alert fatigue among clinical teams, progressively reducing responsiveness to pertinent alerts. Statistical performance and clinical effect are therefore not merely distinct: they may move in opposite directions.

3.5.3. Equity: the most severe limitation

The asymmetry in sensitivity observed among children of sub-Saharan African ancestry, children under 5 years of age and those outside tertiary centres [2] does not arise from an isolated algorithmic defect: it reflects unequal documentation quality across information systems and differing population composition between facilities. A model that performs less well for the populations most exposed produces the opposite of the intended effect. Of the eleven studies analysed, only one performed an equity analysis. This absence reflects not individual negligence on the part of authors, but a publication norm that has not yet incorporated equity as a validation criterion.

3.5.4. Validity of the outcome measure

A critical systematic review notes that all-cause in-hospital mortality, used as the outcome in the derivation cohort of the Phoenix criteria, "may not fully reflect sepsis-associated mortality and could introduce misclassification of outcomes" [35]. An outcome measure based on all-cause mortality will attribute to sepsis deaths arising from other determinants, and vice versa.

3.5.5. Methodological limitations stated explicitly

Several limitations must be stated at the outset, as they condition the scope of the conclusions. The reported performance measures derive from studies whose populations, operational definitions and algorithms differ, which precludes any quantitative aggregation — a prohibition that is itself a finding, as

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

the discussion will establish. It should also be noted that the period covers 2024, the year in which the diagnostic criteria were revised, introducing temporal heterogeneity between earlier and later studies. Finally, since negative studies or those without usable performance data are less likely to be published, a non-quantifiable publication bias persists.

4. Discussion

4.1. Advances of the past five years

First advance. Early detection of pediatric sepsis rests on an established biological foundation. Microvascular phenotypes precede and predict organ failure independently of conventional covariates [25]; glycocalyx biomarkers quantify a causal mechanism whose kinetics precede clinically documented failure [26,27]; proteomic profiles combine pathogen detection, immune activation and endothelial injury [31]. These findings, derived from distinct populations and distinct methods, converge: the phase of dysregulation is detectable before organ failure.
Second advance. Predictive performance depends less on algorithmic refinement than on the quality of covariates and the robustness of validation. Differences between complex and parsimonious models are of the same order of magnitude as between-sample variability [14,21,22], and three quarters of neonatal models have undergone only a single validation [23]. Computational refinement is therefore not the priority lever for improvement; data quality and the repetition of validations are.
Third advance. Implementation, not modelling, constitutes the bottleneck. The dissociation between 94% and 36% sensitivity [2,21] measures the distance separating statistical demonstration from clinical effect. A model whose performance has been validated only on research data has not established its capacity to alter patient outcomes**.**

4.2. Relationship between prediction and therapeutic intervention

Prediction has value only if the intervention it triggers is effective. Recent trials serve as a reminder that such effectiveness is neither universal nor intuitive — a reminder that constitutes a major constraint for the field.

The 2026 Surviving Sepsis Campaign guidelines for children substantially revised fluid resuscitation: administration of 40 to 60 mL/kg in fractionated boluses (10 to 20 mL/kg per bolus) within the first hour is suggested where intensive care is available (conditional recommendation, low certainty of evidence); withholding fluid boluses is formally recommended in children with sepsis without hypotension where intensive care is unavailable (strong recommendation, high certainty of evidence) [36].

The PRoMPT BOLUS trial, involving 9,041 children treated for septic shock, found no significant difference between balanced crystalloids and isotonic saline for the primary outcome (3.4% versus 3.0%; difference 0.4 percentage points; 95% CI −0.5 to 1.3; relative risk 1.10; 95% CI 0.88–1.40), despite a reduction in hyperchloraemia (31.4% versus 49.0%) [37]. The median number of days free of hospitalisation at 28 days was identical (23 days).

An updated meta-analysis of nine studies confirms the absence of any reduction in mortality (RR 0.94; 95% CI 0.83–1.06; $I^2 = 0\%$) or in acute kidney injury (RR 0.84; 95% CI 0.61–1.16; $I^2 = 73.5\%$), with the required information size not reached, leaving these results indeterminate. It does, however, demonstrate a reduction in the use of renal replacement therapy (RR 0.74; 95% CI 0.56–0.96; $I^2 = 23.5\%$) and in hyperchloraemia (RR 0.65; 95% CI 0.61–0.69; $I^2 = 0\%$) [38].

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

Regarding vasopressors, a propensity score-matched cohort study found that administration of norepinephrine was associated with lower mortality than epinephrine in children without known cardiac dysfunction (3.7% versus 0%; risk difference 3.7%; 95% CI 0.2–7.2); this is an observational association, susceptible to residual confounding by indication, and not an established causal relationship [39]. A multicentre prospective observational study of 225 patients reports that epinephrine remained the first-line agent in 70% of cases, followed by norepinephrine in 27% [40].

Regarding corticosteroids, the 2026 guidelines suggest not using intravenous hydrocortisone where haemodynamic stability can be restored by fluid resuscitation and vasoactive agents (conditional recommendation, low certainty of evidence), and consider the evidence insufficient to formulate a recommendation in refractory shock [36].

These findings delineate the framework within which a predictive alert exerts its effect — a narrow one. Anticipating the identification of a patient alters outcome only if the intervention subsequently initiated is effective. This is why demonstration of a clinical effect, rather than isolated statistical performance, constitutes the ultimate criterion of validity for a model.

4.3. Conditions of validity for a deployable model

The synthesis allows six conditions of validity to be stated, ranked according to their necessity and their operational feasibility.

First — Alignment with the 2024 consensus criteria. A model tested on earlier criteria measures a clinical entity that is no longer recognised [5,6].
Second — External validation, and where possible repeated. Three quarters of neonatal models have undergone only a single validation [23]; a model validated in its population of origin has established only its goodness of fit.
Third — Temporal validation. A model tested on data predating the pandemic may show altered performance as a result of changes in epidemiology, practice and documentation.
Fourth — Stratified equity analyses. The asymmetry in performance according to ethnicity, age and type of facility [2] makes equity a validation criterion rather than an optional refinement.
Fifth — Evaluation on real-world data. The dissociation between research and deployment [2,21] establishes that such evaluation does not duplicate the evaluation conducted during development.
Sixth — Demonstration of clinical benefit. None of the studies analysed conducted a randomised trial comparing model-assisted decision-making with conventional clinical decision-making; several authors explicitly acknowledge that this evaluation remains to be performed [17,22]. This central deficiency of the current literature may invalidate the interpretation of reported performance as evidence of efficacy.

4.4. Perspectives

Three directions appear to be priorities, in descending order of feasibility.

First — Incorporation of biomarkers linked to pathophysiological mechanisms. Glycocalyx biomarkers quantify a causal mechanism [26,27], and the neonatal proteomic panel combines pathogen detection, leucocyte recruitment and endothelial injury [31]. Their integration could improve specificity, where models based solely on conventional covariates plateau — subject to prospective validation within a multimodal model, which has not yet been conducted.

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

Second — Stratification by endotype. Identification of a targetable endotype combining aberrant STAT signalling and T-lymphocyte dysregulation [28] opens the way to prediction directed not at the occurrence of sepsis, but at a determined pathophysiological trajectory. This direction remains contingent on the generation of large-scale multi-omic data, which limits its immediate feasibility.

Third — Evaluation of clinical impact, an absolute priority. One review notes that no study has evaluated the clinical impact of predictive technologies for pediatric sepsis [22]. In the absence of a randomised trial comparing model-assisted decision-making with conventional clinical decision-making, the reported performance measures a predictive capacity, not a benefit to the patient.

4.5. Limitations of the present review

This review is based on a search restricted to the period 2021–2026.
No formal instrument for quantitative risk-of-bias assessment was applied, and the reported performance derives from studies that differ in populations, operational definitions and algorithms.
This heterogeneity, far from constituting a mere methodological shortcoming, is itself a finding: it establishes that a quantitative meta-analysis performed on these data would measure heterogeneity of methods rather than the intrinsic performance of the models.

5. Conclusion

Three conclusions emerge, in descending order of strength.
First conclusion. Early detection of pediatric sepsis rests on an established biological foundation. Microvascular phenotypes identified by an unbiased approach precede and predict organ failure independently of conventional clinical variables, with mortality ranging from 1.9% to 17.2% according to the observed profile. In parallel, glycocalyx biomarkers quantify a causal pathophysiological mechanism directly correlated with prognosis. In children, the singularity of the immune response — reflected in an IL-10/TNF-α ratio fifteen times higher than in adults and in the coexistence of hyperinflammation and immunosuppression — makes this window for anticipation both indispensable and particularly narrow [32,33,34].
Second conclusion. Predictive performance depends less on the complexity of the model than on the quality of covariates and the robustness of validation. Differences between complex and parsimonious models are of the same order of magnitude as between-sample variability [14,21,22], and three quarters of neonatal models have undergone only a single validation [23].
Third conclusion. The principal limiting factor now lies in the clinical integration of these tools, rather than in their modelling. The collapse in sensitivity, falling from 94% in the index study to 36% in routine practice [2,21] — with a particularly marked loss of performance among children of sub-Saharan African ancestry, children under 5 years of age and those outside tertiary centres — directly implicates both the developers and the authorities responsible for deployment.
The major gap remains the absence of any evaluation of clinical impact.
No randomised trial has compared a decision based on a predictive model with conventional clinical decision-making [17,22]. Until such a demonstration is produced, the clinical value of predictive models for pediatric sepsis remains a biologically grounded and methodologically unestablished hypothesis.

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

November 2026 • Volume 1 • Number 1-ENGV

Declarations

Author contributions (CRediT): Ahmed HADDADI1; Karima HADDADI2; Souad HADDADI3
1. MD, PhD, Polyclinique Médicale GROUPE AVERROES; Research Unit EA 2694 Public Health, Epidemiology and Quality of Care, Medical University, Lille, France — a.haddadi@averroes-emj.com
2. MD, MSc, Hôpital Antoine Béclère — karima.haddadi@aphp.fr
3. PhD, Polyclinique Médicale GROUPE AVERROES — souad.h@averroes-emj.com
Conflicts of interest: the authors declare no conflicts of interest relating to this work.
Funding: this work received no funding.

All authors approved the submitted version. Ethical approval: narrative review without collection of individual data; no ethical approval was required.
The data analysed derive exclusively from the published literature.

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

References

1. Umbrella review for pediatric sepsis and robust prediction potential worldwide. Pediatr Discov. 2024. PMCID: PMC13612974.

2. Evaluation of a commercial pediatric sepsis early detection alert in the emergency department. JAMA Pediatr. 2026. PMID: 42832216.

3. The burden and contemporary epidemiology of sepsis in children. Lancet Child Adolesc Health. 2024;8(9):683-694.

4. Global, regional, and national sepsis incidence and mortality, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Glob Health. 2025.

5. Schlapbach LJ, Watson RS, Sorce LR, et al. International consensus criteria for pediatric sepsis and septic shock. JAMA. 2024;331(8):665-674.

6. Sanchez-Pinto LN, Bennett TD, DeWitt PE, et al. Development and validation of the Phoenix criteria for pediatric sepsis and septic shock. JAMA. 2024;331(8):675-686.

7. National estimates of pediatric sepsis in US hospitals using clinical data. Pediatrics. 2026. PMID: 41865411.

8. Etiology, prevalence, and mortality of sepsis among children under five years in Africa: a systematic review and meta-analysis. BMC Infect Dis. 2026;26:13350.

9. The burden of neonatal sepsis and its risk factors in Africa: a systematic review and meta-analysis. BMC Public Health. 2025;25:22076.

10. Early-onset sepsis in preterm neonates of 22-28 weeks' gestation: an international cohort study. J Pediatr. 2026.

11. Multidrug-resistant sepsis in special newborn care units in five district hospitals in India: a prospective cohort study. Lancet Glob Health. 2025;13(5).

12. Beyond early- and late-onset neonatal sepsis definitions. Pediatr Infect Dis J. 2024;43(12).

13. An interpretable machine learning model using routine clinical indicators for early pediatric sepsis risk stratification: a multicenter retrospective cohort study. Front Cell Infect Microbiol. 2026;16:1933201.

14. Development and external validation of age-specific predictive models for sepsis in critically ill neonates and children using multidimensional hematological parameters. Crit Care. 2026;30:6261.

15. Construction and efficacy evaluation of a model for early diagnosis of pediatric sepsis based on LASSO-logistic regression. 2025. PMID: 40933706.

16. Predicting pediatric sepsis and mortality using wearable device data and machine learning in Bangladesh. 2025. PMID: 41252746.

17. A machine learning model for prediction of early-onset neonatal sepsis in low-income and middle-income countries. BMJ Paediatr Open. 2026;10(1):e003561.

18. The Phoenix criteria and other severity scores in pediatric sepsis. JAMA Netw Open. 2026;9(6).

19. External validation of the Phoenix sepsis score in a paediatric intensive care unit in Saudi Arabia. Sci Rep. 2026;16:48807.

20. Aligning prediction models with clinical information needs: infant sepsis case study. JAMIA Open. 2025;8(2):ooaf015.

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved

21. Derivation and validation of predictive models for early pediatric sepsis. JAMA Pediatr. 2025. doi:10.1001/jamapediatrics.2025.3167

22. A scoping review on pediatric sepsis prediction technologies in healthcare. npj Digit Med. 2024;7:342.

23. Clinical prediction models to diagnose neonatal sepsis in low-income and middle-income countries: a scoping review. 2025. PMID: 40204466.

24. Artificial intelligence in pediatric intensive care units: current applications in sepsis management. World J Pediatr. 2026.

25. Microvascular phenotypes in pediatric sepsis identified by machine learning. Crit Care. 2026;30.

26. Association between serum endothelial glycocalyx biomarkers and prognosis in paediatric sepsis. BMC Infect Dis. 2026;26:12688.

27. Biomarkers of glycocalyx injury and endothelial activation in sepsis: systematic review and meta-analysis. J Intensive Care Med. 2023;38(10).

28. Aberrant STAT signaling and T cell dysregulation define a targetable pediatric sepsis endotype. J Clin Invest. 2026;136(14):e202867.

29. Identification and transcriptomic assessment of latent profile pediatric septic shock phenotypes. Crit Care. 2024;28:241.

30. Predictive value of dynamic plasma biomarkers for clinical outcomes in pediatric sepsis. 2025. PMID: 41636191.

31. Serum proteomics in newborns reveals a multi-pathway molecular panel for neonatal sepsis. Sci Rep. 2026;16:78660.

32. Immunomodulation in pediatric sepsis: a narrative review. J Clin Med. 2025;14(9):2983.

33. Molecular mechanisms underlying inflammation in early-onset neonatal sepsis. 2025. PMC12347463.

34. Sepsis-related immunosuppression: a comprehensive review of current research breakthroughs. Crit Care. 2026;30.

35. The Phoenix criteria and other severity scores in identifying pediatric sepsis. PMC13231296.

36. Surviving Sepsis Campaign international guidelines for the management of sepsis and septic shock in children 2026. Intensive Care Med. 2026. doi:10.1007/s00134-026-08360-2

37. Balanced fluid or 0.9% saline in children treated for septic shock (PRoMPT BOLUS). N Engl J Med. 2026. doi:10.1056/NEJMoa2601969

38. Balanced crystalloids versus normal saline for fluid resuscitation in pediatric septic shock: an updated systematic review and meta-analysis. 2026. PMC13719638.

39. Epinephrine vs norepinephrine as initial treatment in children with septic shock. JAMA Netw Open. 2025;8(4).

40. Management and outcome of paediatric septic shock requiring vasoactive therapy: a multicentre prospective observational study. 2026. PMID: 43012742.

41. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

42. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-58.

Copyright © 2026 by the Society of Averroes-EMJ All Rights Reserved