Research Article
E-ISSN : 3048-1368
Volume 49, Issue No. 02. 2026
P-ISSN : 3048-1376

The International Journal of Medical Science and Health Research



EFFECTIVENESS OF PREDICTIVE ANALYTICS ALGORITHMS FOR EARLY SEPSIS DETECTION IN THE INTENSIVE CARE UNIT ON TIME TO INTERVENTION AND MORTALITY : A SYSTEMATIC REVIEW AND META-ANALYSIS OF RANDOMIZED CONTROLLED TRIALS AND PRIMARY STUDIES

Yudha Viantoro[1]*, Fifa Yuniarmi[2]

[1] Panggul Primary Health Care Centre, Indonesia

[2] Panggul Regional General Hospital, Indonesia

*Corresponding E-mail: yudhaviantoro01@gmail.com

Article History :

Received date : 2026/07/15

Revised date : 2026/08/21

Accepted date : 2026/09/12

Published date : 2026/10/30

Copyright: *© 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (BY NC) license (https://creativecommons.org/licenses/by-nc/4.0/).*

E-ISSN :

ISSN 3048-1368

P-ISSN

ISSN 3048-1376

ABSTRACT

Background: Sepsis remains a leading cause of death in the intensive care unit (ICU), and every hour of delay in antimicrobial therapy is associated with higher mortality. Predictive analytics algorithms — rule-based electronic surveillance and machine-learning (ML) models embedded in the electronic health record (EHR) — promise earlier recognition, but whether they shorten time to intervention and reduce mortality in critically ill patients remains uncertain.

Objective: To synthesise evidence from randomised controlled trials (RCTs) and primary studies on the effectiveness of predictive analytics for early sepsis detection in ICU settings on time to intervention, mortality, length of stay (LOS) and diagnostic performance.

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

Methods: The study strictly adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines. Eligible designs were RCTs, comparative studies, cohorts and case series conducted in ICUs or in mixed hospital settings that included ICU patients; systematic reviews were excluded from synthesis and used for discussion only. Risk of bias was assessed with RoB 2, ROBINS-I and PROBAST. Random-effects meta-analyses (DerSimonian–Laird with modified Hartung–Knapp confidence intervals [CIs] and 95% prediction intervals [PIs]) were performed, and certainty of evidence was rated with GRADE.

Results: 26 studies (27 reports) were included: 4 RCTs, 10 comparative cohort/pre–post studies and 12 diagnostic validation cohorts. Across nine studies (30,144 patients), predictive analytics was associated with a non-significant reduction in mortality (odds ratio [OR] 0.71, 95% CI 0.49–1.02; I^2 = 75%; PI 0.34–1.49), with a consistent direction for the risk ratio (0.66, 0.46–0.94) and the absolute risk difference (-3.8 percentage points, -6.8 to -0.8). The benefit was confined to ML-based/predictive systems (OR 0.60, 0.42–0.84; k = 5) rather than rule-based alerts (OR 0.99, 0.42–2.37; k = 4; P for interaction = 0.089) and disappeared when studies at high/serious risk of bias were excluded (OR 1.01, 0.44–2.32). Time to first antibiotic was shorter by a mean of 1.83 h (95% CI -4.57 to 0.90; k = 5; I^2 = 88%) and ICU LOS by 1.10 days (-2.92 to 0.72), whereas hospital LOS was unchanged (0.27 days, -3.61 to 4.14). Guideline-concordant care tended to improve (OR 1.43, 0.85–2.41). Pooled discrimination across 15 validation datasets was AUROC 0.864 (0.816–0.901; PI 0.60–0.96), similar in ICU-only (0.854) and mixed settings (0.879), with very high heterogeneity (I^2 = 99.7%).

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

Conclusions: Predictive analytics for sepsis, particularly ML-based early warning systems, achieve good discrimination and show a consistent signal towards earlier treatment and lower mortality in critically ill patients; however, the certainty of evidence is low to very low because most data derive from non-randomised designs with serious risk of bias and substantial heterogeneity. Adequately powered, pragmatic multicentre RCTs that measure the complete alert-to-action pathway, alert burden and local calibration are required before universal adoption.

Keywords: sepsis; septic shock; machine learning; artificial intelligence; predictive analytics; early warning system; clinical decision support; intensive care unit; time to antibiotics; mortality; meta-analysis

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

INTRODUCTION

Background

Sepsis is one of the most consequential syndromes in contemporary medicine. The Global Burden of Disease analysis estimated 48.9 million incident cases and 11.0 million sepsis-related deaths in 2017, accounting for 19.7% of all global deaths, and the updated 2021 estimate rose to approximately 166 million cases and 21.4 million sepsis-related deaths, representing 31.5% of global mortality after the COVID-19 pandemic reversed earlier gains. The burden is disproportionately concentrated in low- and middle-income regions, including South-East Asia.[1,2]

The intensive care unit (ICU) carries the highest concentration of sepsis morbidity. A meta-analysis of population-level data estimated an incidence of ICU-treated sepsis of 58 per 100,000 person-years, with 41.9% of patients dying before hospital discharge, and the Intensive Care over Nations audit identified sepsis during the ICU stay in 29.5% of 10,069 patients, with a hospital mortality of 35.3%. Importantly, approximately one in four ICU sepsis cases is acquired during the ICU stay itself, creating a window in which continuous surveillance could theoretically detect deterioration before overt organ failure. In Asian ICUs, the MOSAICS II study found sepsis in 22.4% of admissions, a hospital mortality of 32.6%, compliance with the 1-hour bundle of only 21.5%, and a 2.5-fold increase in the adjusted odds of death when antibiotics were delayed beyond 3 hours.[3–6]

The Third International Consensus Definitions (Sepsis-3) define sepsis as life-threatening organ dysfunction caused by a dysregulated host response to infection, operationalised as an acute increase in the Sequential Organ Failure Assessment (SOFA) score of at least 2 points. Among ICU encounters with suspected infection, SOFA discriminated in-hospital mortality better (AUROC 0.74) than systemic inflammatory response syndrome (SIRS) criteria (0.64) or the quick SOFA (0.66), and paediatric sepsis is now identified with the Phoenix Sepsis Score. Because these criteria are confirmatory rather than anticipatory, clinicians frequently recognise sepsis only after organ dysfunction is established.[7–9]

Time is a modifiable determinant of outcome. In 49,331 patients treated under New York State mandates, each hour of delay in completing the 3-hour bundle increased the risk-adjusted odds of in-hospital death by 4%, and each hour of delay in antibiotic administration had a similar effect; comparable hourly penalties were reported in Northern California (OR 1.09 per hour), for 1-year mortality (10% increase per hour), in septic shock (35% increased risk per hour) and in a planned secondary analysis of a cluster-randomised trial conducted in 40 German ICUs (0.42% absolute increase in 28-day mortality per hour of delay). The Surviving Sepsis Campaign (SSC) therefore recommends performance-improvement programmes that include screening of acutely ill, high-risk

patients and antimicrobial administration within 1 hour for septic shock or probable sepsis, recommendations that were maintained in the 2026 update.[10–16]

Predictive analytics refers to the automated, continuous computation of an individual's risk of sepsis from routinely collected EHR data, with a resulting alert or decision-support output intended to trigger earlier clinical action (Figure 1). The spectrum ranges from rule-based electronic surveillance ("sniffers" applying SIRS or organ-dysfunction logic) to machine-learning (ML) models such as the Targeted Real-time Early Warning Score (TREWScore), InSight, the Artificial Intelligence Sepsis Expert (AISE), COMPOSER, NAVOY Sepsis and the proprietary Epic Sepsis Model (ESM), which integrate high-dimensional, time-varying physiological and laboratory signals.[17–22]

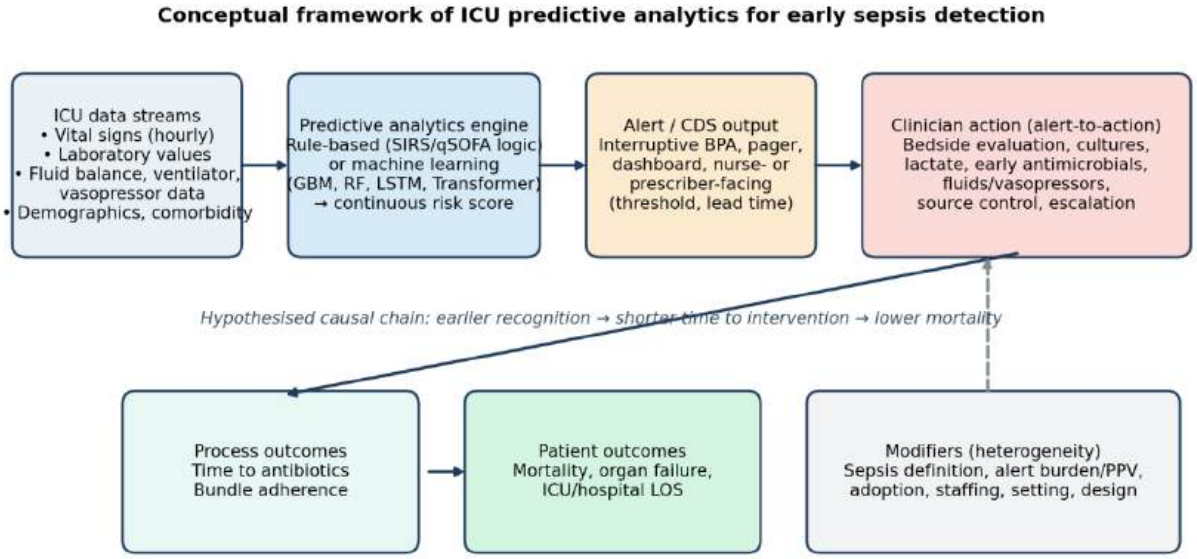


Figure 1. Conceptual framework of ICU predictive analytics for early sepsis detection. Continuous EHR data streams feed a rule-based or machine-learning engine that generates a risk score and an alert; the clinical benefit depends on the alert-to-action step (bedside evaluation, cultures, early antimicrobials, haemodynamic support), which in turn is hypothesised to shorten time to intervention and reduce mortality. Effect modifiers (sepsis definition, alert burden and positive predictive value, clinician adoption, staffing, setting and study design) explain heterogeneity across studies.[15,23]

Previous evidence syntheses have focused predominantly on discrimination rather than on patient outcomes. A diagnostic-accuracy meta-analysis of 28 papers (130 models) reported AUROCs ranging from 0.68 to 0.99 in the ICU but identified only three clinical implementations; a systematic review of 22 ICU studies precluded meta-analysis because of heterogeneity; and a 2025 methodological review of 91 studies showed that the median AUROC fell to 0.783 under full-window external validation. Outcome-focused reviews of automated alerting reported a pooled

mortality risk ratio of 0.71 overall but a non-significant effect in the ICU (RR 0.90), a risk ratio of 0.81 in emergency departments, and a Cochrane review of three ICU RCTs graded the evidence as very low certainty.[24–29]

Research gap

Four gaps justify the present review. First, no published synthesis has jointly quantified process outcomes (time to intervention), patient outcomes (mortality and LOS) and discrimination specifically for the ICU population. Second, the most recent outcome meta-analysis that reported an ICU-specific estimate predates the prospective TREWS evaluation, the Epic Sepsis Model implementation study, the NAVOY randomised validation and the 2026 multicentre deployment studies. Third, external validation of a widely deployed proprietary model demonstrated poor discrimination (AUROC 0.63), falling to 0.47 when predictions made after clinical recognition were excluded, raising concern that discrimination reported in development cohorts may not translate into earlier treatment. Fourth, the comparative effectiveness of ML-based versus rule-based systems on hard outcomes has not been examined with contemporary random-effects methods appropriate for few, heterogeneous studies.[22,27,29–31]

Novelty

This review is, to our knowledge, the first to (i) integrate RCTs, comparative cohorts and diagnostic validation cohorts published within the last 15 years (2011–2026) in a single ICU-centred framework; (ii) quantify the alert-to-action pathway through time-to-antibiotic and bundle-adherence outcomes alongside mortality; (iii) stratify effects by algorithm class (ML-based versus rule-based) and report 95% prediction intervals using the modified Hartung–Knapp approach recommended for small meta-analyses; and (iv) apply three design-specific risk-of-bias instruments (RoB 2, ROBINS-I and PROBAST) together with GRADE, while incorporating evidence published up to 2026.[32–36]

Objectives

General objective. To evaluate the effectiveness of predictive analytics algorithms for early sepsis detection in ICU settings on time to intervention and mortality. **Specific objectives.** (i) To estimate the effect of predictive analytics, compared with usual care, on all-cause mortality; (ii) to estimate its effect on time to first antibiotic administration and adherence to guideline-concordant sepsis care; (iii) to estimate its effect on ICU and hospital LOS; (iv) to summarise the discrimination (AUROC), sensitivity, specificity and lead time of ICU sepsis prediction algorithms; and (v) to explore whether effects differ by study design, algorithm class and clinical setting.[15,16]

Hypotheses

H1 (primary): compared with usual care, predictive analytics reduces all-cause mortality in patients at risk of sepsis in the ICU (H0: OR = 1). **H2:** predictive analytics shortens the time to first

antibiotic administration (H0: mean difference = 0). **H3:** ML-based systems are more effective than rule-based alerts in reducing mortality (H0: no subgroup difference). **H4:** ICU sepsis prediction algorithms achieve good discrimination (pooled AUROC ≥ 0.80) in external validation.[10,27]

Significance and expected benefits

Clinically, the findings inform intensivists and hospital leaders deciding whether to adopt, configure or de-implement sepsis alerts, and identify the conditions (algorithm class, alert design, adoption) under which benefit is most plausible. Institutionally, the review provides implementation guidance for health systems with expanding EHR infrastructure, including Indonesian referral networks linking primary health care centres with regional hospitals, where delayed recognition and low bundle compliance remain common. Scientifically, the review delineates priorities for pragmatic trials and for reporting standards such as TRIPOD+AI.[6,15,37]

METHODS

Protocol and reporting standards

This systematic review and meta-analysis was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement, supplemented by the TRIPOD-SRMA checklist for the prediction-model components. The protocol, including eligibility criteria, outcomes and analysis plan, was developed a priori.[32,38]

Eligibility criteria

Eligibility was defined using the PICOS framework. Sepsis was accepted as defined by the primary studies (Sepsis-1/2, Sepsis-3, Centers for Disease Control and Prevention Adult Sepsis Event or validated diagnostic codes), and paediatric ICU studies were eligible because the Phoenix criteria now align paediatric and adult definitions conceptually. Only RCTs, comparative studies (including controlled before–after and interrupted time-series designs), cohorts (including diagnostic/prognostic validation cohorts) and case series were eligible. Systematic reviews and meta-analyses were not included in the characteristics tables or quantitative synthesis but were used for comparison in the Discussion.[7,9]

Table 1. Eligibility criteria according to the PICOS framework.

Element	Inclusion criteria	Exclusion criteria
Population	Critically ill adults or children in ICUs (medical, surgical, mixed, paediatric) or hospital-wide/mixed settings in which ICU patients formed part of the monitored population; patients at risk of, or with suspected, sepsis/septic shock	Emergency department-only, ward-only or pre-hospital populations; neonatal ICUs; non-human or simulated data
Intervention /	Automated predictive analytics for sepsis	Manual paper screening tools;

index test	embedded in or applied to EHR data: rule-based electronic surveillance/alerts and machine-learning or statistical prediction models generating real-time risk scores or alerts	prediction of mortality or LOS only (prognostic models not aimed at sepsis detection); reminder systems not involving sepsis detection
Comparator	Usual care, pre-implementation period, silent (non-alerting) operation, non-confirmed alerts, or reference-standard diagnosis (for accuracy studies)	No comparator and no extractable accuracy data
Outcomes	Primary: all-cause mortality; time to first antibiotic. Secondary: ICU and hospital LOS; bundle/guideline-concordant care; AUROC, sensitivity, specificity; lead time	Outcomes not extractable or reported only graphically without numerical data
Study design	RCTs, comparative studies (controlled before–after, interrupted time series, propensity-matched cohorts), prospective/retrospective cohorts including validation cohorts, case series	Systematic reviews/meta-analyses (used for discussion only), narrative reviews, editorials, protocols, conference abstracts, preprints, case reports
Time and language	Published January 2011 to 25 October 2026 (maximum 15 years); English; peer-reviewed full text	Publications before 2011

EHR, electronic health record; ICU, intensive care unit; LOS, length of stay; RCT, randomised controlled trial.

Search Strategy

The keywords used for this research based PICO :

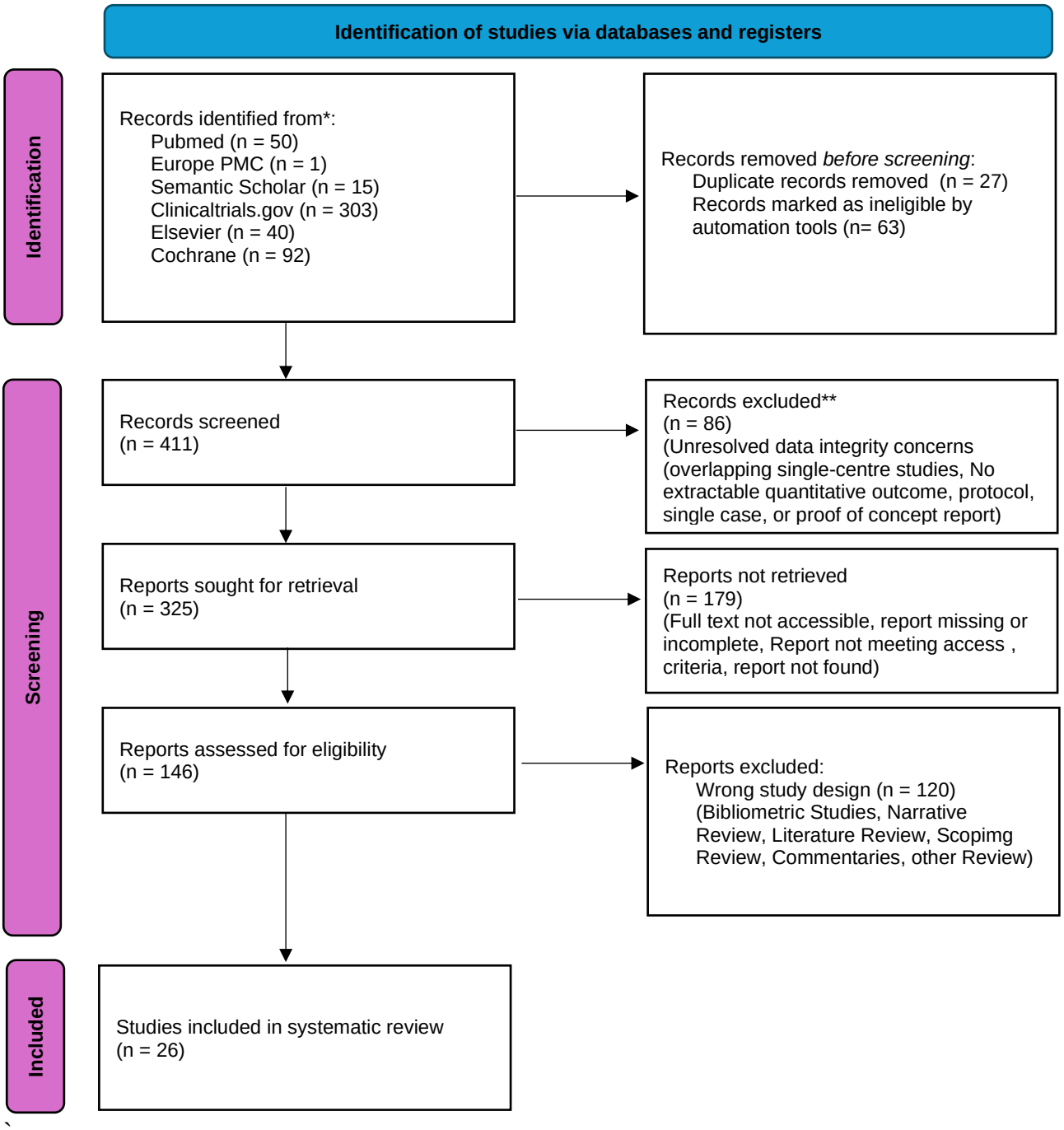
PICO Element	Keyword 1	Keyword 2	Keyword 3
P (Population)	Sepsis	Intensive Care Unit	Critically Ill Patients
I (Intervention)	Predictive Analytics Algorithms	Machine Learning	Early Sepsis Detection
C (Comparison)	Standard Care	Conventional Sepsis Screening	Usual Clinical Practice

O (Outcome)	Mortality	Time to Intervention	Time to Antibiotic Administration

The Boolean MeSH keywords inputted on databases for this research are: *("Sepsis" OR "Intensive Care Unit" OR "Critically Ill Patients") AND ("Predictive Analytics Algorithms" OR "Machine Learning" OR "Early Sepsis Detection") AND ("Standard Care" OR "Conventional Sepsis Screening" OR "Usual Clinical Practice") AND ("Mortality" OR "Time to Intervention" OR "Time to Antibiotic Administration")*

Article Search Strategy

Database	**Keywords**	**Hits**
Pubmed	*("Sepsis" OR "Intensive Care Unit" OR "Critically Ill Patients") AND ("Predictive Analytics Algorithms" OR "Machine Learning" OR "Early Sepsis Detection") AND ("Standard Care" OR "Conventional Sepsis Screening" OR "Usual Clinical Practice") AND ("Mortality" OR "Time to Intervention" OR "Time to Antibiotic Administration")*	50
Europe PMC	*("Sepsis" OR "Intensive Care Unit" OR "Critically Ill Patients") AND ("Predictive Analytics Algorithms" OR "Machine Learning" OR "Early Sepsis Detection") AND ("Standard Care" OR "Conventional Sepsis Screening" OR "Usual Clinical Practice") AND ("Mortality" OR "Time to Intervention" OR "Time to Antibiotic Administration")*	1
Semantic Scholar	*("Sepsis" OR "Intensive Care Unit" OR "Critically Ill Patients") AND ("Predictive Analytics Algorithms" OR "Machine Learning" OR "Early Sepsis Detection") AND ("Standard Care" OR "Conventional Sepsis Screening" OR "Usual Clinical Practice") AND ("Mortality" OR "Time to Intervention" OR "Time to Antibiotic Administration")*	15
Clinicaltrials.gov	*(“Sepsis” OR “Intensive Care Unit” OR “Critically Ill Patients”) AND (“Predictive Analytics Algorithms” OR “Machine Learning” OR “Early Sepsis Detection”) AND (“Mortality” OR “Time to Intervention” OR “Time to Antibiotic Administration”)*	303
Elsevier	*(“Sepsis” OR “Intensive Care Unit” OR “Critically Ill Patients”) AND (“Predictive Analytics Algorithms” OR “Machine Learning” OR “Early Sepsis Detection”) AND (“Mortality” OR “Time to Intervention” OR “Time to Antibiotic Administration”)*	40
Cochrane	*(“Sepsis” OR “Intensive Care Unit” OR “Critically Ill Patients”) AND (“Predictive Analytics Algorithms” OR “Machine Learning” OR “Early Sepsis Detection”) AND (“Mortality” OR “Time to Intervention” OR “Time to Antibiotic Administration”)*	92



Article search flowchart

Selection process

Records were de-duplicated by digital object identifier (DOI) and title. Titles and abstracts were screened against the eligibility criteria using a pre-specified keyword-assisted filter (requiring sepsis terms combined with prediction, detection, alert, surveillance or decision-support terms, and excluding reviews, editorials, prognostic-only models and neonatal populations), followed by verification of retained and borderline records by the reviewers. Full texts were assessed independently by two reviewers (YV and FY), and disagreements were resolved by discussion. Reasons for full-text exclusion were recorded.[32]

Data collection and data items

Data were extracted into a standardised form capturing bibliographic details, country, design, setting, population, sample size, sepsis definition, algorithm class and name, input variables, alert modality and recipients, comparator, outcome definitions and time points, effect estimates with measures of dispersion, discrimination metrics, thresholds, prediction horizon and lead time. When studies reported medians with interquartile ranges, means and standard deviations (SDs) were estimated with the methods of Wan and Luo; standard errors were converted to SDs using group sizes; counts were derived from reported percentages and denominators; asymmetric CIs were symmetrised on the analysis scale; and SEs of AUROC were derived from reported CIs, from fold-level SDs, or from the numbers of events and non-events (Hanley–McNeil approximation), as recommended for prediction-model meta-analyses.[39–41]

Outcomes

The primary outcomes were all-cause mortality (in-hospital, ICU or 28-day; the longest reported in-hospital horizon was preferred) and time from alert, recognition or arrival to first antibiotic administration or order. Secondary outcomes were ICU LOS, hospital LOS, adherence to guideline-concordant sepsis care (bundle completion, order-set use or recommended interventions), discrimination (AUROC), sensitivity and specificity at the implemented threshold, and lead time before sepsis onset or treatment.[10,15]

Risk-of-bias assessment

Risk of bias was assessed at the outcome level by two reviewers using the revised Cochrane tool for randomised trials (RoB 2; five domains), ROBINS-I for non-randomised studies of interventions (seven domains, including confounding and selection of reported results), and PROBAST for prediction-model validation studies (participants and data sources, predictors, outcome, analysis), complemented by the PROBAST+AI signalling considerations for ML models.[33–35,42]

Risk of bias

Among the four RCTs, three were judged to have some concerns and one high risk of bias, chiefly because of open-label delivery of alerts, unblinding through the alert itself and the absence of pre-registered analysis plans. Among the ten non-randomised studies, two were at moderate, seven at serious and one at critical risk of bias, driven by confounding inherent to uncontrolled before–after designs and secular trends in sepsis care. Among the 15 validation datasets, four were at low, three at unclear and eight at high risk of bias; the analysis domain (internal-only validation, small numbers of events or thresholds derived on the test data) and the outcome domain (diagnostic-code-based sepsis labels) were the most frequent sources of concern.[33–35,42]

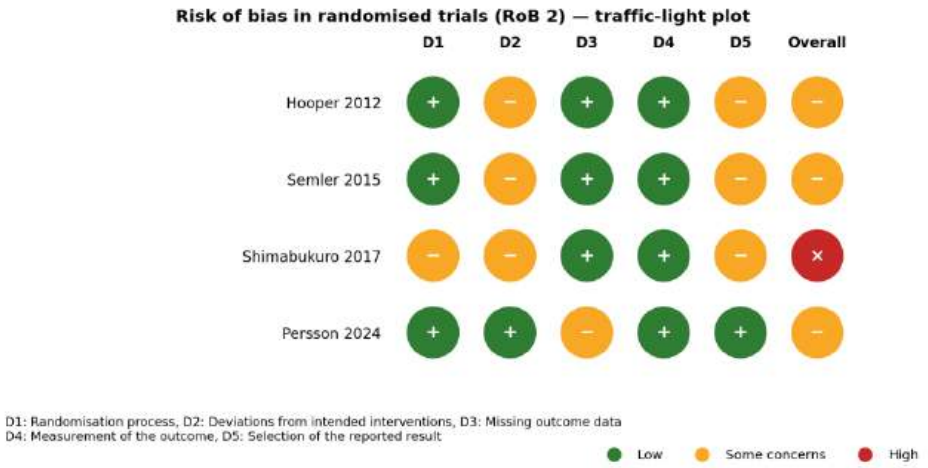


Figure 2. Risk-of-bias traffic-light plot for randomised controlled trials (RoB 2). D1, randomisation process; D2, deviations from intended interventions; D3, missing outcome data; D4, measurement of the outcome; D5, selection of the reported result.[33]

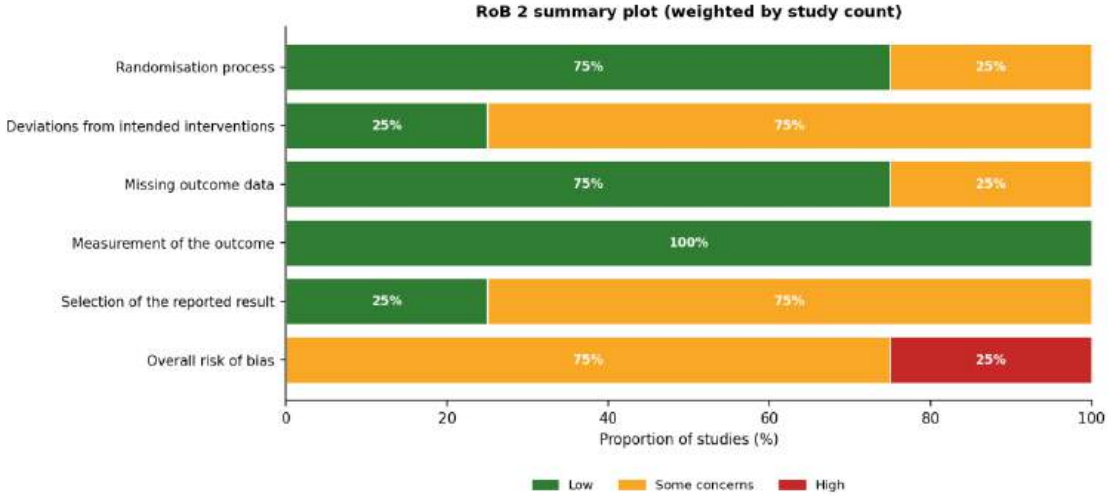


Figure 3. RoB 2 summary plot showing the proportion of trials at low risk, some concerns and high risk of bias in each domain.[33]

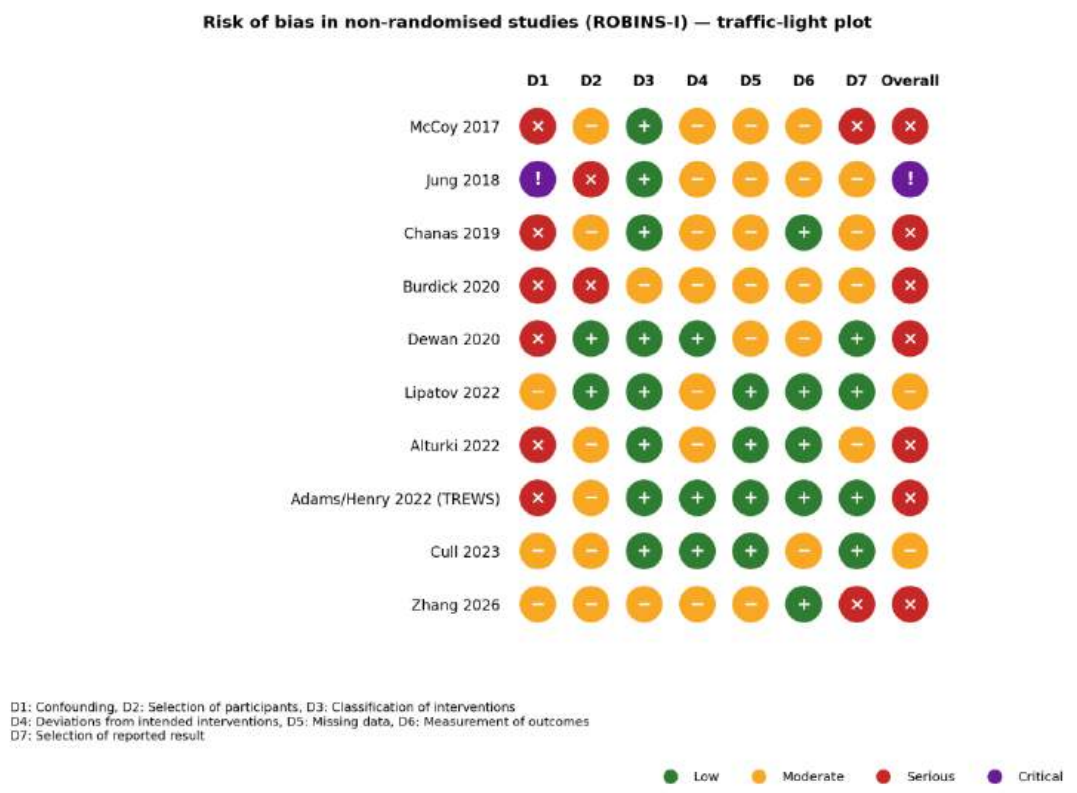


Figure 4. Risk-of-bias traffic-light plot for non-randomised studies of interventions (ROBINS-I).[34]

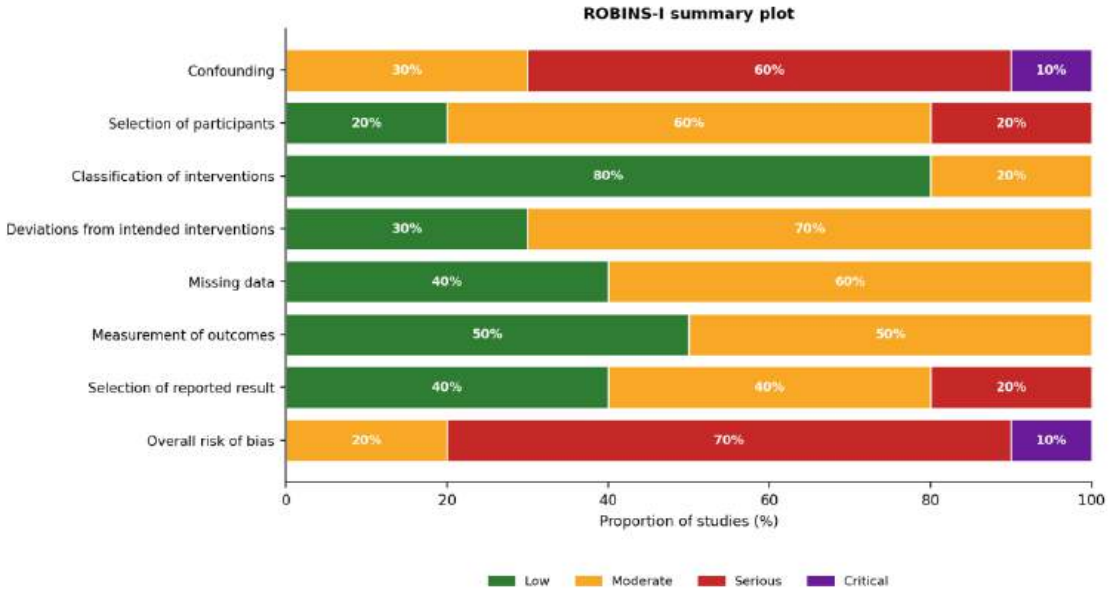


Figure 5. ROBINS-I summary plot by domain. Confounding was the dominant source of serious or critical risk of bias.[34]



Figure 6. Risk-of-bias traffic-light plot for prediction-model validation studies (PROBAST). D1, participants and data sources; D2, predictors; D3, outcome; D4, analysis.[35]

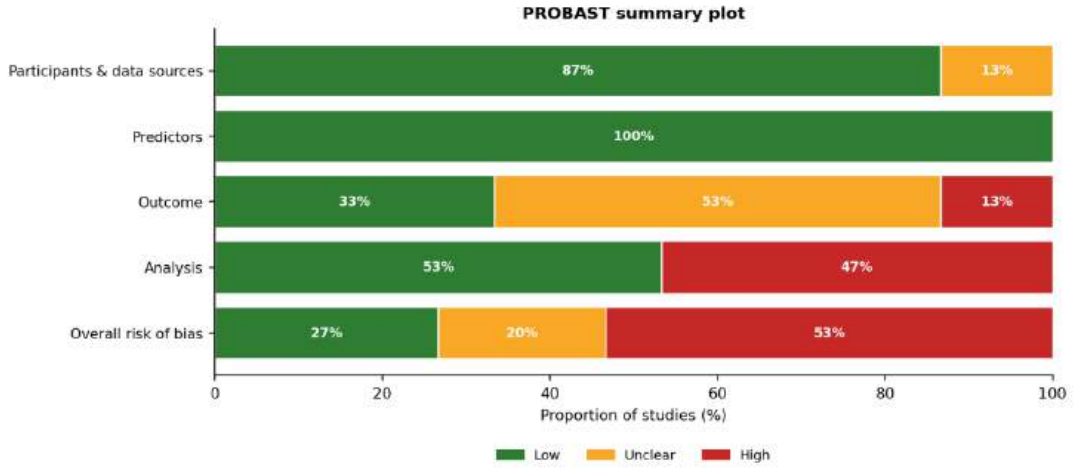


Figure 7. PROBAST summary plot by domain. The analysis and outcome domains accounted for most high or unclear judgements.[35,42]

Table 2. RoB 2 domain-level judgements for randomised controlled trials.[33]

Study	D1	D2	D3	D4	D5	Overall	Principal concern
Hooper 2012	Low	Some concerns	Low	Low	Some concerns	Some concerns	Open-label; no pre-registered protocol (2009 trial)
Semler 2015	Low	Some concerns	Low	Low	Some concerns	Some concerns	Open-label; tool used for orders in only 28% of cases
Shimabukuro 2017	Some concerns	Some concerns	Low	Low	Some concerns	High	Allocation revealed upon alerts; vendor-affiliated analysis; small sample
Persson 2024	Low	Low	Some concerns	Low	Low	Some concerns	Only 85 of the standard-care arm analysable for accuracy

D1, randomisation process; D2, deviations from intended interventions; D3, missing outcome data; D4, measurement of the outcome; D5, selection of the reported result. Green, low; yellow, some concerns; red, high.

Table 3 ROBINS-I domain-level judgements for non-randomised studies of interventions.[34]

Study	D1	D2	D3	D4	D5	D6	D7	Overall	Principal concern
McCoy 2017	Serious	Moderate	Low	Moderate	Moderate	Moderate	Serious	Serious	Uncontrolled pre–post; secular trends; vendor co-authors

Study	D1	D2	D3	D4	D5	D6	D7	Overall	Principal concern
Jung 2018	Critical	Serious	Low	Moderate	Moderate	Moderate	Moderate	Critical	n = 7 vs 23; no adjustment
Chanas 2019	Serious	Moderate	Low	Moderate	Moderate	Low	Moderate	Serious	Pre-period size not reported
Burdick 2020	Serious	Serious	Moderate	Moderate	Moderate	Moderate	Moderate	Serious	Baseline sometimes first month post-go-live
Dewan 2020	Serious	Low	Low	Low	Moderate	Moderate	Low	Serious	Pre–post; chart-review outcomes
Lipatov 2022	Moderate	Low	Low	Moderate	Low	Low	Low	Moderate	ITS adjusted for pre-trends
Alturki 2022	Serious	Moderate	Low	Moderate	Low	Low	Moderate	Serious	Retrospective pre–post; crude comparison
Adams/Henry 2022 (TREWS)	Serious	Moderate	Low	Low	Low	Low	Low	Serious	Exposure = clinician confirmation (confounding by behaviour)
Cull 2023	Moderate	Moderate	Low	Low	Low	Moderate	Low	Moderate	Multivariable adjustment; ICD-based sepsis

Study	D1	D2	D3	D4	D5	D6	D7	Overall	Principal concern
Zhang 2026	Moderate	Moderate	Moderate	Moderate	Moderate	Low	Serious	Serious	Implementation subgroup sizes not reported

D1, confounding; D2, selection of participants; D3, classification of interventions; D4, deviations from intended interventions; D5, missing data; D6, measurement of outcomes; D7, selection of the reported result. Purple, critical.

Table 4. PROBAST domain-level judgements for prediction-model validation datasets.[35,42]

Study	Participants	Predictors	Outcome	Analysis	Overall	Principal concern
Henry 2015	Low	Low	Unclear	High	High	Internal only; complex model
Desautels 2016	Low	Low	Unclear	High	High	Internal CV only
McCoy 2017	Unclear	Low	Unclear	High	High	Limited reporting of analysis
Nemati 2018	Low	Low	Unclear	Low	Unclear	External validation
Mao 2018	Low	Low	Unclear	High	High	Internal; ICD-based labels
Barton 2019	Low	Low	Unclear	Low	Unclear	External test set
Burdick 2020 (BMC)	Low	Low	High	Low	High	ICD-based outcome
Wong 2021	Low	Low	Low	Low	Low	Full external validation
Shashikumar 2021	Low	Low	Low	Low	Low	External/temporal validation
Wang 2021	Unclear	Low	Unclear	High	High	Single centre; internal only

Study	Participants	Predictors	Outcome	Analysis	Overall	Principal concern
Cull 2023	Low	Low	High	High	High	Cut-point derived and tested on same data
Moor 2023	Low	Low	Low	Low	Low	International external validation
Kim 2023	Low	Low	Unclear	Low	Unclear	Temporal validation
Persson 2024	Low	Low	Low	High	High	Only 20 events
Dutta 2026	Low	Low	Low	Low	Low	Prospective silent deployment

Green, low; yellow, unclear; red, high risk of bias.

Effect measures and synthesis methods

The odds ratio (OR) was the primary summary measure for mortality because adjusted estimates were reported as ORs in several non-randomised studies; risk ratios (RRs) and risk differences (RDs) were calculated as secondary measures. Continuous outcomes were pooled as mean differences (MDs; hours or days) and, for time to antibiotics, additionally as standardised mean differences (Hedges' g). Discrimination was pooled on the logit-AUROC scale and back-transformed. Random-effects models with the DerSimonian–Laird estimator of between-study variance (τ^2) were fitted, and 95% CIs were computed with the modified Hartung–Knapp–Sidik–Jonkman (mKH) method using a t distribution with $k - 1$ degrees of freedom; conventional Wald-type CIs are reported for comparison. Heterogeneity was quantified with Cochran's Q, I^2 and τ^2, and 95% prediction intervals (PIs) were calculated with $k - 2$ degrees of freedom.[36,43–45]

Subgroup analyses, sensitivity analyses and meta-regression

Pre-specified subgroup analyses compared RCTs versus non-randomised studies, ML-based/predictive versus rule-based systems, and ICU-only versus mixed settings for mortality, and ICU-only versus mixed settings and internal versus external validation for discrimination; subgroup differences were tested with a between-group Q statistic. Sensitivity analyses comprised a common-effect model, Wald-type CIs, substitution of adjusted for crude estimates, exclusion of studies with indirect interventions or derived counts, exclusion of studies at high/serious risk of bias, restriction to adults, leave-one-out analysis and cumulative meta-analysis by publication year. Exploratory

random-effects meta-regression examined publication year and baseline (control-group) mortality; subgroup credibility was interpreted with GRADE guidance.[43,46]

Reporting bias

Small-study effects were explored with contour-enhanced funnel plots, Egger's regression test and the trim-and-fill method (L0 estimator). Because fewer than ten studies contributed to most analyses, these assessments were interpreted as exploratory.[47,48]

Certainty of evidence

Certainty was rated with GRADE for each outcome across the domains of risk of bias, inconsistency, indirectness, imprecision and publication bias; bodies of evidence from non-randomised studies started at low certainty, and the GRADE approach for test accuracy was applied to discrimination outcomes.[46,49,50]

Software and ethics

Analyses were performed in Python 3 (NumPy, SciPy and Matplotlib) using scripts implementing the DerSimonian–Laird, modified Hartung–Knapp, Egger and trim-and-fill estimators; the extraction dataset and code are available from the corresponding author. Ethical approval was not required because only published, de-identified aggregate data were analysed.[36,43]

RESULTS

Characteristics of included studies

Four RCTs were included: two single-centre trials of rule-based electronic tools in medical and surgical ICUs at a tertiary academic centre in the United States, one open-label RCT of an ML algorithm (InSight) in two medical–surgical ICUs, and one blinded randomised clinical validation of the NAVOY Sepsis algorithm in a Swedish ICU. Together the trials randomised 1,295 patients.[21,55–57]

Ten non-randomised comparative studies evaluated implementation of predictive analytics: four ML or proprietary predictive systems in mixed hospital settings including ICUs (InSight in a community hospital and across nine US hospitals, the TREWS programme across five hospitals, and the Epic Sepsis Model), one ML trajectory model implemented in ICUs with propensity-score matching, and five rule-based systems in surgical, medical and paediatric ICUs. Study populations ranged from 30 to 17,758 analysed patients.[23,30,58–66]

Twelve diagnostic validation cohorts (plus the accuracy components of two implementation studies and one RCT) reported discrimination of sepsis prediction algorithms, nine derived exclusively from ICU data and six from mixed settings that included ICU patients. Algorithms comprised gradient-boosted trees, random forests, Weibull–Cox proportional-hazards models, deep recurrent and attention-based networks, conformal deep-learning models and proprietary EHR-vendor models; sepsis was defined by Sepsis-3 in most recent cohorts and by diagnostic codes or

surveillance definitions in others. Most evidence originated from the United States, with additional cohorts from Sweden, Saudi Arabia, China, South Korea, the Netherlands and Switzerland. No eligible case series reporting comparative outcomes was identified.[17–20,22,67–73]

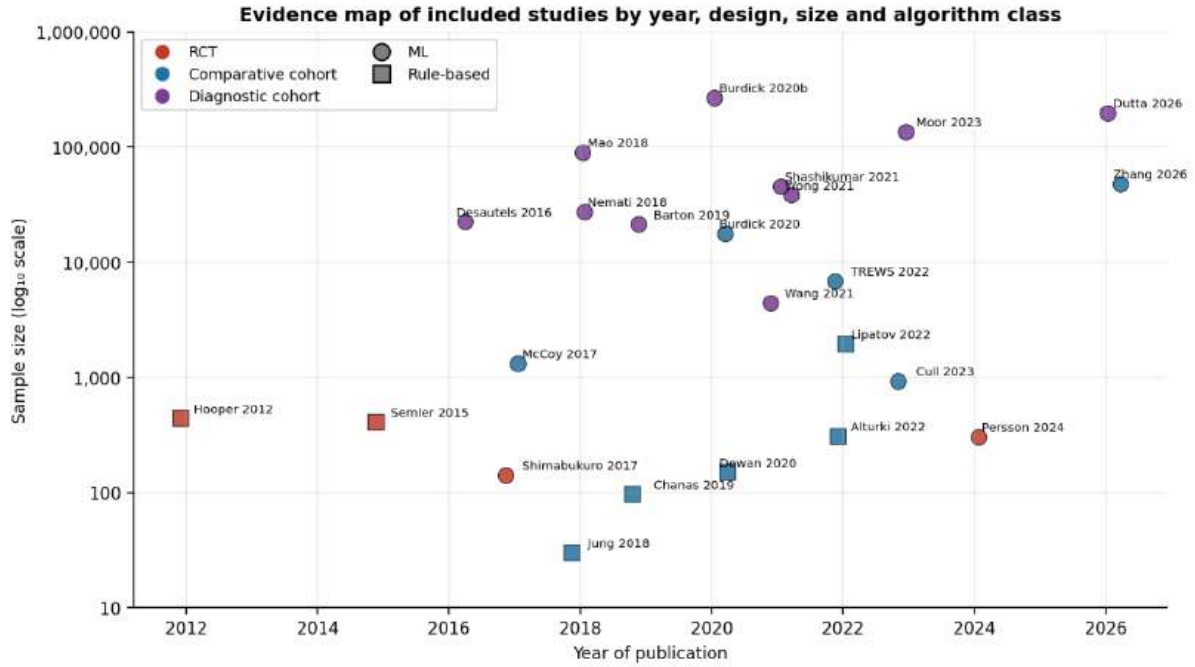


Figure 8. Evidence map of the 26 included studies by year of publication, sample size (log scale), design (colour) and algorithm class (marker shape). Henry 2015 is not plotted because the cohort size was not reported in a form allowing comparison. Diagnostic cohorts are large retrospective EHR datasets, whereas randomised evidence is limited to four small-to-moderate single-centre trials.[17,61]

Table 5. Characteristics of included randomised and non-randomised implementation studies (k = 14).[21,23,30,55–66]

Study	Country	Design	Setting and population	N (intervention / control)	Algorithm (class; name)	Alert modality and comparator	Outcomes extracted	Risk of bias
Hooper 2012	USA	RCT, single centre	Medical ICU; adults meeting modified SIRS criteria	220 / 222	Rule-based: real-time "Listening Application" (modified SIRS)	Pager + EMR flag to physicians vs usual care (no alert)	Mortality; time to new antibiotic; ICU and hospital LOS	Some concerns
Semler 2015	USA	Pragmatic RCT, single centre	Medical and surgical ICUs; adults with sepsis	218 / 189	Rule-based CDS: electronic sepsis evaluation and management tool (logic rules, guideline display, order entry)	Tool added to electronic sepsis alerting vs alerting alone	ICU mortality; 6-h bundle completion (HR)	Some concerns
Shimabukuro 2017	USA	Open-label RCT, 2 ICUs	Medical–surgical ICUs; adults	67 / 75	ML (InSight; gradient	MLA alerts + existing detector	In-hospital mortality;	High

					boosting on vital signs)	vs existing detector only	hospital and ICU LOS	
Persson 2024	Sweden	Blinded randomised clinical validation	General ICU; adults	304 randomised (85 analysable for accuracy)	ML (NAVOY Sepsis; 4-h vital signs, blood gas, laboratory values)	Active alerts vs silent predictions	Sensitivity, specificity, AUROC (3 h ahead); LOS (no difference)	Some concerns
McCoy 2017	USA	Prospective QI before–after	ED, ICU and wards; adults with ≥ 2 SIRS criteria	921 / 407	ML (InSight)	MLA alerts vs twice-daily SIRS screening	Sepsis-related in-hospital mortality; LOS; readmission	Serious
Jung 2018	USA	Pre–post cohort	Surgical ICU; confirmed sepsis	7 / 23	Rule-based visual sepsis screening score in bedside surveillance display	Display with vs without screening score	Time from positive score to antibiotics; LOS	Critical
Chanas 2019	USA	Pre–post cohort	Surgical ICU; septic shock	97 post (pre NR)	Rule-based best-practice advisory (cultures +	Advisory with team/pharmacist notification vs none	Time to antibiotics (narrative only)	Serious

					vasopressor)			
Burdick 2020	USA	Prospective multicentre before–after (9 hospitals)	ED/inpatient/ICU; adults with ≥ 2 SIRS criteria	14,166 / 3,592	ML (InSight)	MLA period vs baseline period	In-hospital mortality; LOS; readmission	Serious
Dewan 2020	USA	Prospective pre–post cohort	Paediatric ICU; severe sepsis	80 / 70	Rule-based CDS (temperature + culture → nurse perfusion screen)	CDS period vs pre-CDS	Timely recommended interventions; alert accuracy	Serious
Lipatov 2022	USA	Before–after with interrupted time series	Medical ICU and ED; severe sepsis/septic shock	1,096 / 854	Rule-based sepsis "sniffer" + decision support and completion feedback	Post vs pre implementation	Hospital mortality (adjusted OR); bundle compliance	Moderate
Alturki 2022	Saudi Arabia	Retrospective pre–post	Paediatric ICU; children ≤ 14 years with sepsis	161 / 147	Rule-based electronic alert system (SIRS/sepsis)	Post-EAS vs pre-EAS	Mortality; PICU LOS; time to fluids and antibiotics	Serious

Adams 2022; Henry 2022 (TREWS)	USA	Prospective multisite cohort (5 hospitals)	ED, wards and ICUs; 6,877 sepsis patients alerted before antibiotics (590,736 monitored)	Confirmed vs not confirmed within 3 h	ML (TREWS)	Alert confirmed by provider ≤ 3 h vs not confirmed	Adjusted mortality (RR/RD); time to first antibiotic order	Serious
Cull 2023	USA	Before–after	Adult inpatients (38–50% ICU admission); ESM ≥ 5 with sepsis, no prior antibiotics	628 / 304	Proprietary predictive model (Epic Sepsis Model v1)	Active alerts vs silent background scoring	Mortality (crude, adjusted OR); time to antibiotics; LOS; order-set use	Moderate
Zhang 2026	China (+ MIMIC-III, eICU)	Multicentre cohort with propensity-matched implementation	ICU; Sepsis-3 adults (47,936 total)	Implementation subgroup NR	ML ensemble trajectory model (dynamic variability)	Trajectory-guided intervention vs standard care	28-day mortality (RD); ICU LOS; ventilation; antibiotic timeliness	Serious

Risk of bias: RoB 2 for RCTs; ROBINS-I for non-randomised studies. CDS, clinical decision support; EAS, electronic alert system; ED, emergency department; EMR, electronic medical record; ESM, Epic Sepsis Model; HR, hazard ratio; LOS, length of stay; MLA, machine-learning algorithm; NR, not reported; PICU, paediatric intensive care unit; QI, quality improvement; RD, risk difference; SIRS, systemic inflammatory response syndrome; TREWS, Targeted Real-time Early Warning System.

Table 6. Characteristics and discrimination of included sepsis prediction-model validation cohorts (k = 15 datasets).[17–22,58,60,67–73]

Study (model)	Country; data source	Setting	Validation design	Sample (septic)	Algorithm	Reference standard	Horizon / lead time	AUROC (95% CI)	PROBAST
Henry 2015 (TREWScore)	USA; MIMIC-II	ICU	Retrospective development/internal validation	MIMIC-II ICU cohort	Supervised time-to-event score (TREWScore)	Septic shock	Median 28.2 h before shock	0.83 (0.81–0.85)	High
Desautels 2016 (InSight)	USA; MIMIC-III	ICU	Retrospective, internal cross-validation	22,583 stays (11.3% sepsis)	Ensemble ML on vital signs, GCS, age	Sepsis-3	0–4 h	0.880 (SD 0.006)	High
McCoy 2017 (InSight)	USA; community hospital	Mixed incl. ICU	Retrospective local external evaluation	1,665 encounters	Gradient-boosted ML	Sepsis-3	At onset	0.91 (0.90–0.92)	High
Nemati 2018 (AISE)	USA; Emory ICUs + MIMIC-III	ICU	Development + external validation	27,527 (2,375 septic); ~42,000 validation	Modified Weibull–Cox model (65 features)	Sepsis-3	4–12 h ahead	0.83–0.85	Unclear
Mao 2018 (InSight)	USA; UCSF + BIDMC	Mixed incl. ICU	Retrospective multicentre; transfer learning	90,353 UCSF encounters (684,443 total)	Gradient tree boosting (6 vital signs)	Sepsis / severe sepsis / shock	0–4 h	0.92 (0.90–0.93)	High

Study (model)	Country; data source	Setting	Validation design	Sample (septic)	Algorithm	Reference standard	Horizon / lead time	AUROC (95% CI)	PROBAST
Barton 2019	USA; UCSF → BIDMC	ICU	External (train UCSF, test BIDMC)	21,507 (1,024 septic)	Gradient-boosted ensemble (6 vital signs)	Sepsis onset	0–48 h	0.890	Unclear
Burdick 2020 (BMC)	USA; 461 hospitals + community hospital	Mixed incl. ICU	Retrospective; external validation	270,438 analysed	Gradient-boosted ML	Severe sepsis (ICD-based)	0–48 h	0.948 (0.942–0.954)	High
Wong 2021 (ESM v1)	USA; Michigan Medicine	Mixed incl. ICU	Retrospective external validation	38,455 hospitalisations (2,552 sepsis)	Proprietary penalised regression (Epic)	CDC/ICD composite	Hospitalisation level	0.63 (0.62–0.64)	Low
Shashikumar 2021 (COMPOSER)	USA; two health systems	ICU	External and temporal validation	45,812 external ICU (7,913 septic)	Deep learning with conformal prediction	Sepsis-3	Median 12.2 h before antibiotics	0.925	Low
Wang 2021 (RF)	China; Zhengzhou	ICU	Internal split (4:1)	4,449 infected patients	Random forest (55 features)	Sepsis (Sepsis-3)	NR	0.88 (±0.04)	High

Study (model)	Country; data source	Setting	Validation design	Sample (septic)	Algorithm	Reference standard	Horizon / lead time	AUROC (95% CI)	PROBAST
Cull 2023 (ESM v1)	USA; academic centre	Mixed incl. ICU	Local cut-point derivation	11,512 encounters (1,171 sepsis)	Proprietary (Epic)	Diagnosis codes	Encounter level	0.834	High
Moor 2023 (deep learning)	USA, Netherlands, Switzerland; 4 ICU databases	ICU	Internal + international external	136,478 admissions (25,694 sepsis)	Deep learning (time-series)	Sepsis-3 (hourly)	Median 3.7 h (80% sensitivity)	0.761 (0.746–0.770) external	Low
Kim 2023 (DeepSEPS)	South Korea; single centre	ICU	Temporal validation (2020–2021)	Real-world ICU EMR	Deep learning (vital signs, laboratory)	Sepsis-3	Multiple time points	0.789 (0.786–0.792)	Unclear
Persson 2024 (NAVOY)	Sweden; Malmö	ICU	Prospective randomised (silent arm)	85 (20 septic)	ML (NAVOY Sepsis)	Sepsis-3	3 h	0.80	High
Dutta 2026 (ESM v2)	USA; 9 hospitals	Mixed incl. ICU	Prospective silent deployment	198,494 encounters (5,832 Sepsis-3)	Locally trained gradient-boosted trees (Epic v2)	Sepsis-3 / SEP-1 / ASE	Median 3.4 h	0.89 (0.89–0.90)	Low

ASE, Adult Sepsis Event; BIDMC, Beth Israel Deaconess Medical Center; CDC, Centers for Disease Control and Prevention; GCS, Glasgow Coma Scale; ICD, International Classification of Diseases; MIMIC, Medical Information Mart for Intensive Care; NR, not reported; RF, random forest; SEP-1, Severe Sepsis and Septic Shock Management Bundle; UCSF, University of California San Francisco.

Primary outcome: all-cause mortality

Nine studies including 30,144 patients (three RCTs and six non-randomised studies) contributed to the mortality analysis. Predictive analytics was associated with lower odds of death, but the conservative mKH interval included the null: OR 0.71 (95% CI 0.49–1.02), P = 0.063; the conventional DerSimonian–Laird Wald interval excluded it (0.56–0.91, P = 0.006). Heterogeneity was substantial (I^2 = 75%, τ^2 = 0.083; Q = 31.53, P < 0.001), and the 95% PI (0.34–1.49) indicates that a future setting could experience either a substantial benefit or no benefit.[30,55–60,65,66]

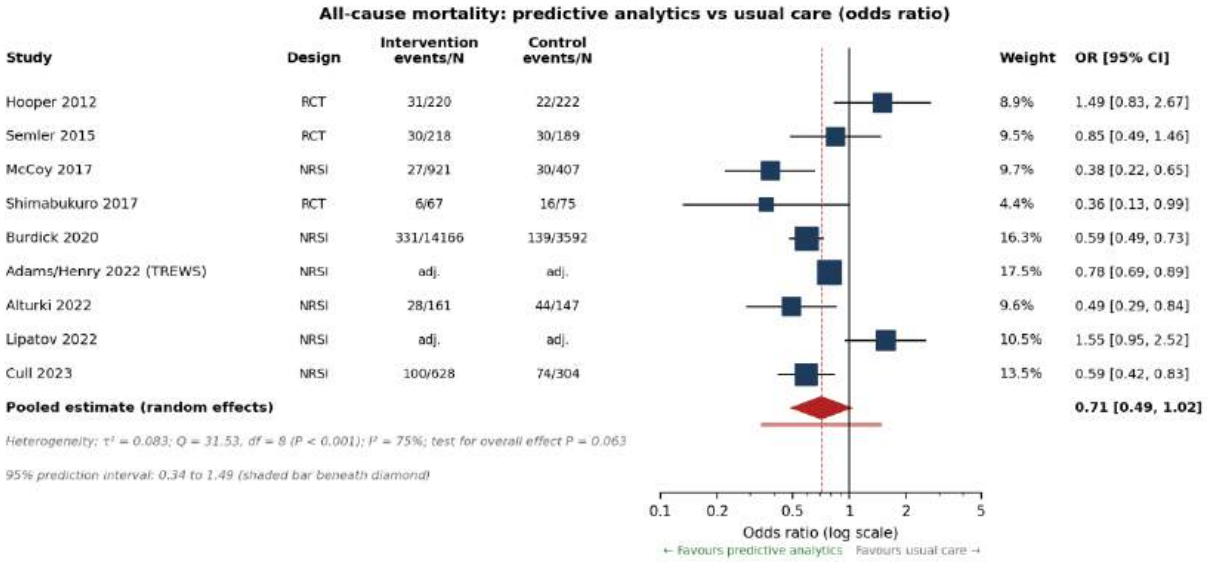


Figure 9. Forest plot of all-cause mortality (odds ratio) comparing predictive analytics with usual care. Squares are proportional to random-effects weights; the red diamond shows the pooled estimate with the modified Hartung–Knapp 95% CI and the shaded bar the 95% prediction interval. For Lipatov 2022 and TREWS only adjusted estimates were available (adj.).[30,55,65]

The direction of effect was consistent across effect measures. In the seven studies with extractable counts, the pooled RR was 0.66 (95% CI 0.46–0.94) (P = 0.029; I^2 = 62%), and across nine studies the absolute RD was -3.8 percentage points (95% CI -6.8 to -0.8; P = 0.019; I^2 = 72%), corresponding to a number needed to treat of approximately 26 patients exposed to predictive analytics to prevent one death. The RD analysis additionally incorporated the adjusted absolute reductions reported by the TREWS programme (3.3 percentage points) and by the ICU trajectory-model implementation (5.7 percentage points).[30,61]

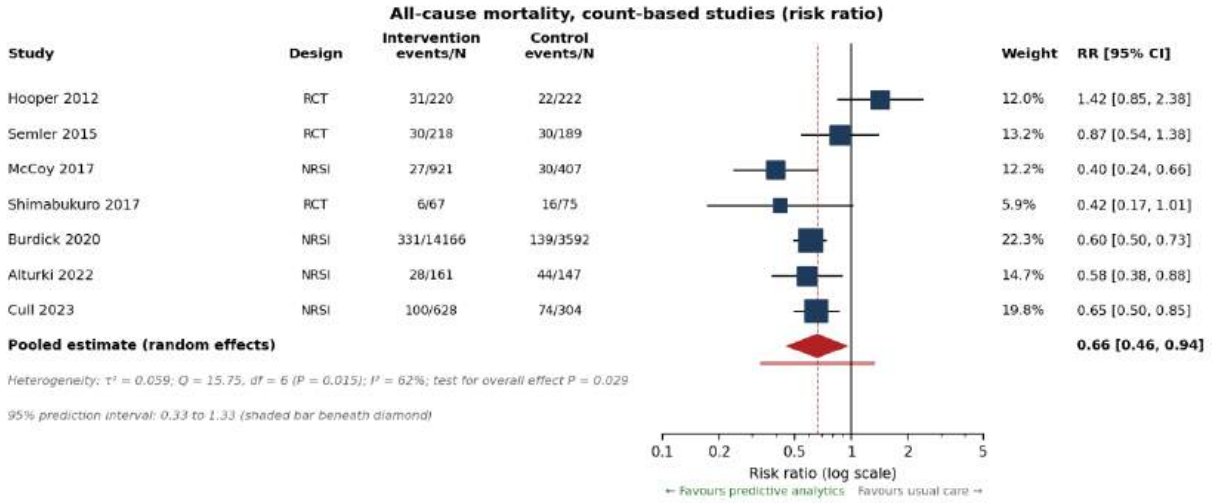


Figure 10. Forest plot of all-cause mortality expressed as risk ratios (studies with extractable event counts).[57,58]

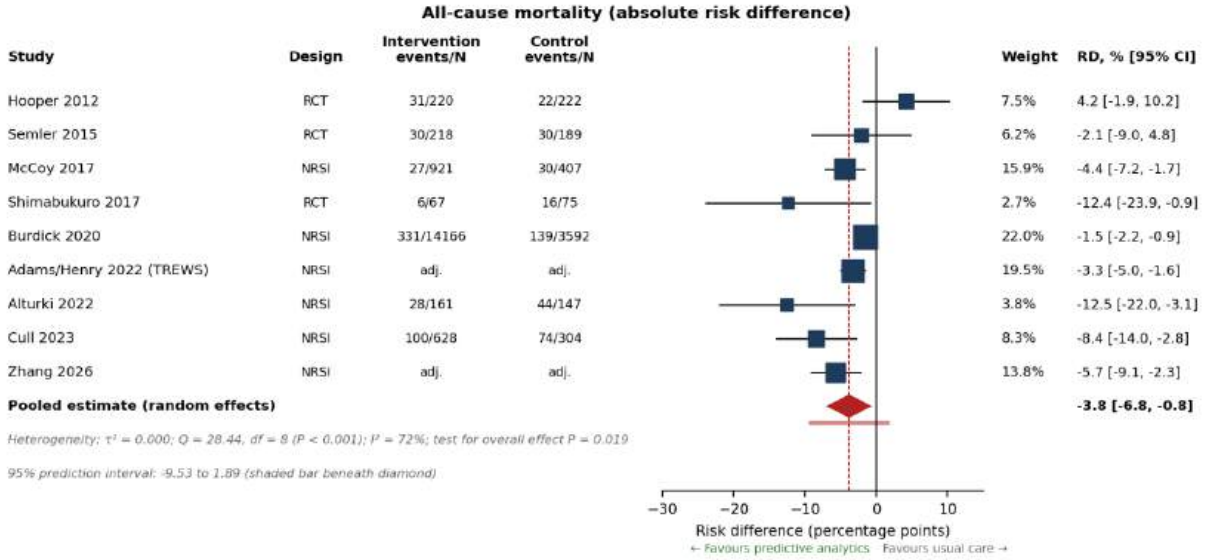


Figure 11. Forest plot of the absolute risk difference in mortality (percentage points). Negative values favour predictive analytics.[30,61]

Subgroup analyses

The effect differed by algorithm class. ML-based or predictive systems were associated with a significant reduction in mortality (OR 0.60 (95% CI 0.42–0.84); k = 5; I^2 = 69%), whereas rule-based alerts were not (OR 0.99 (95% CI 0.42–2.37); k = 4; I^2 = 75%); the test for subgroup difference approached significance (Q = 2.88, P = 0.089). The pooled OR among RCTs (0.85 (95% CI 0.16–4.43)) was imprecise and did not differ from that among non-randomised studies (0.66 (95% CI 0.42–1.05); P = 0.509). Estimates were similar in ICU-only (0.73 (95% CI 0.28–1.90)) and mixed settings (0.69 (95% CI 0.39–1.22); P = 0.872).[30,55–60,65,66]

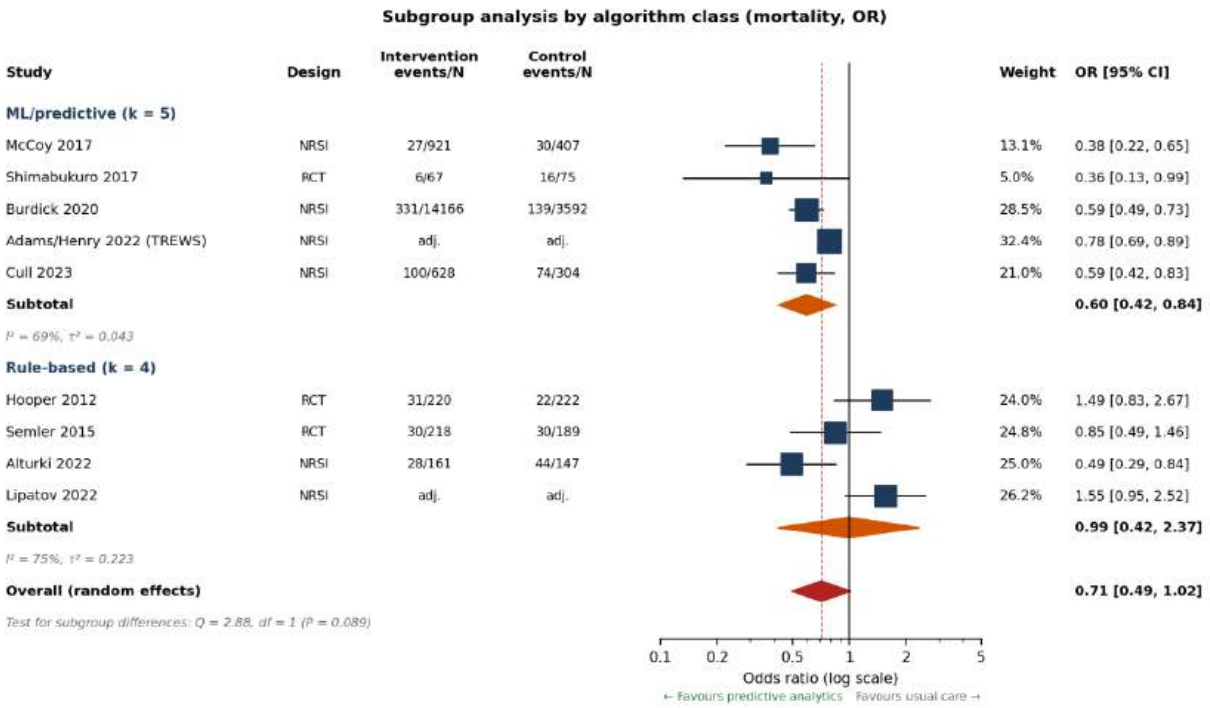


Figure 12. Subgroup analysis of mortality by algorithm class (machine-learning/predictive versus rule-based).[46]

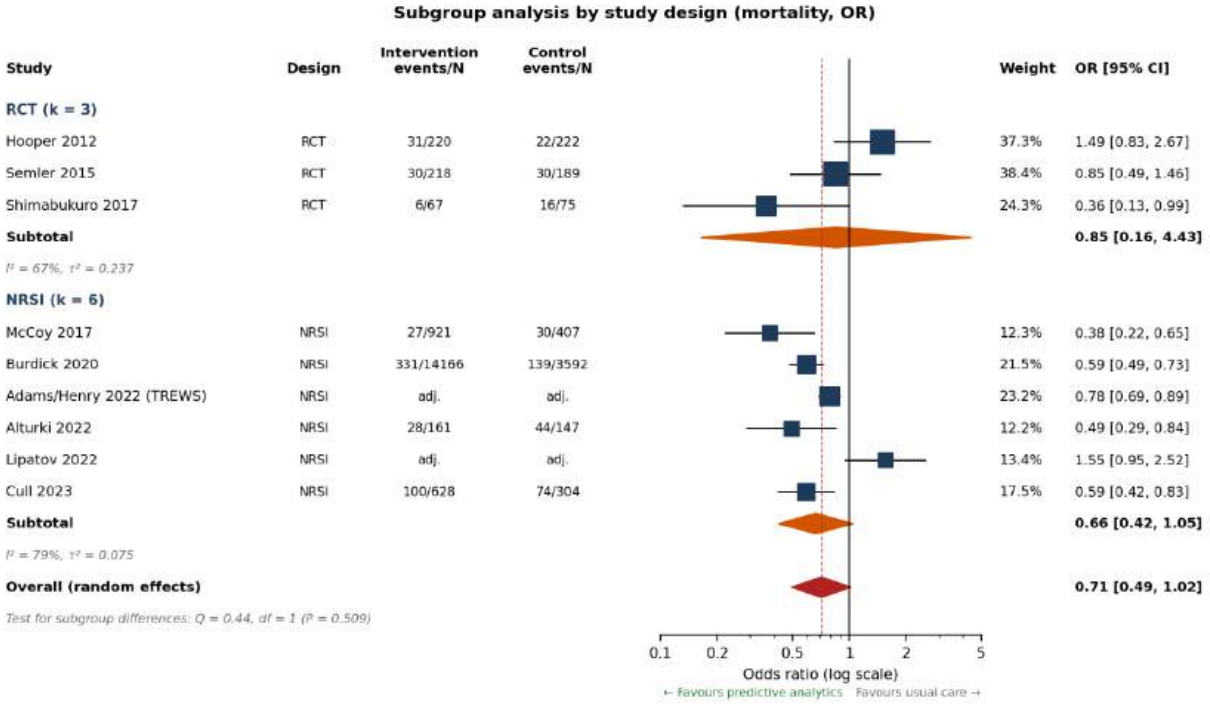


Figure 13. Subgroup analysis of mortality by study design (randomised controlled trials versus non-randomised studies).[46]

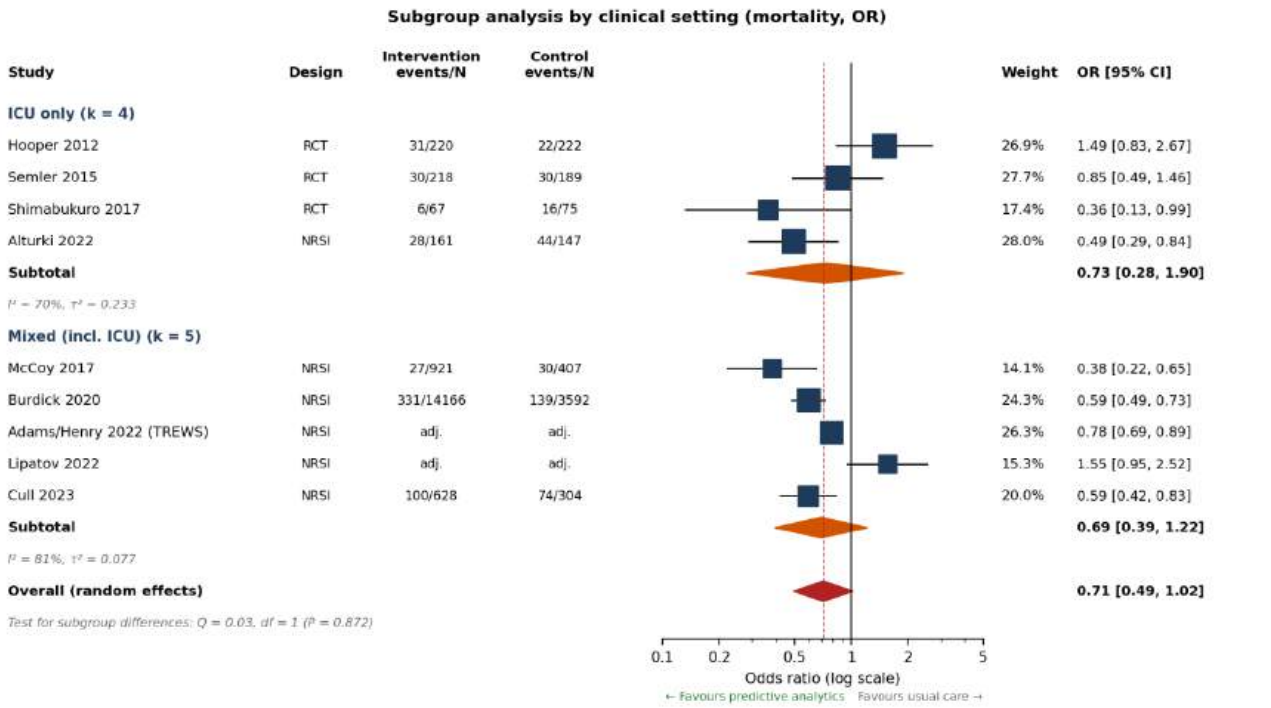


Figure 14. Subgroup analysis of mortality by clinical setting (ICU-only versus mixed hospital settings including ICU patients).[46]

Sensitivity analyses, leave-one-out and cumulative meta-analysis

Results were robust to the choice of model and to most analytical decisions. Substituting the adjusted OR for the crude estimate in the Epic Sepsis Model study (OR 0.71), excluding the indirect-intervention trial (0.70), excluding TREWS (0.70) or excluding derived event counts (0.67) did not materially change the point estimate. However, restricting the analysis to studies at low-to-moderate risk of bias or with some concerns (k = 4) abolished the effect (OR 1.01 (95% CI 0.44–2.32)), indicating that the apparent benefit is driven largely by studies at serious risk of bias. In leave-one-out analysis the pooled OR ranged from 0.65 (omitting Lipatov 2022) to 0.76 (omitting McCoy 2017), with the interval excluding the null only when either of two rule-based studies (Hooper 2012 or Lipatov 2022) was omitted. Cumulative meta-analysis showed that the estimate stabilised around 0.66–0.73 after the addition of the first large ML implementation studies in 2020–2022.[30,56,58–60,65]

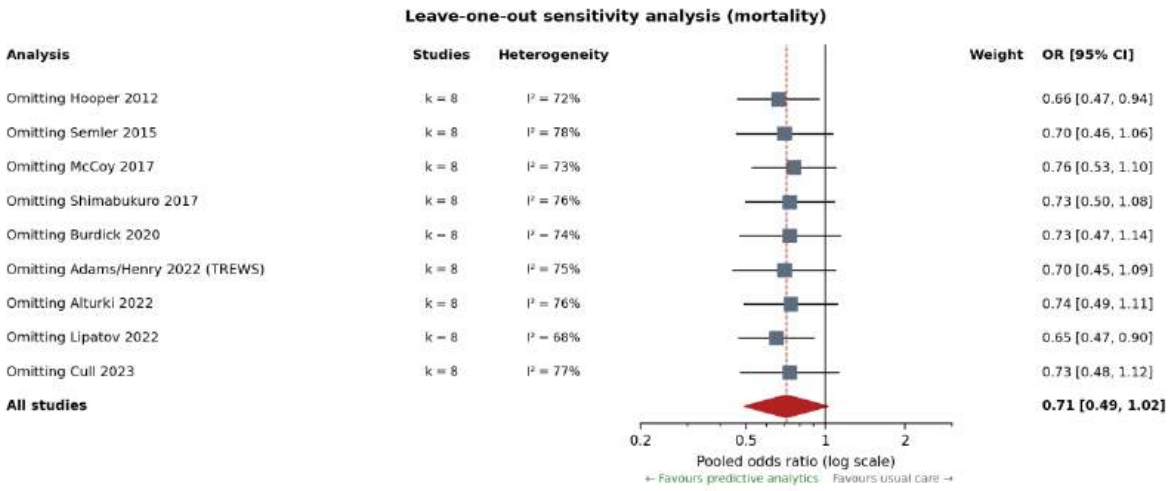


Figure 15. Leave-one-out sensitivity analysis of mortality. Each row shows the pooled odds ratio after omitting the indicated study.[43]

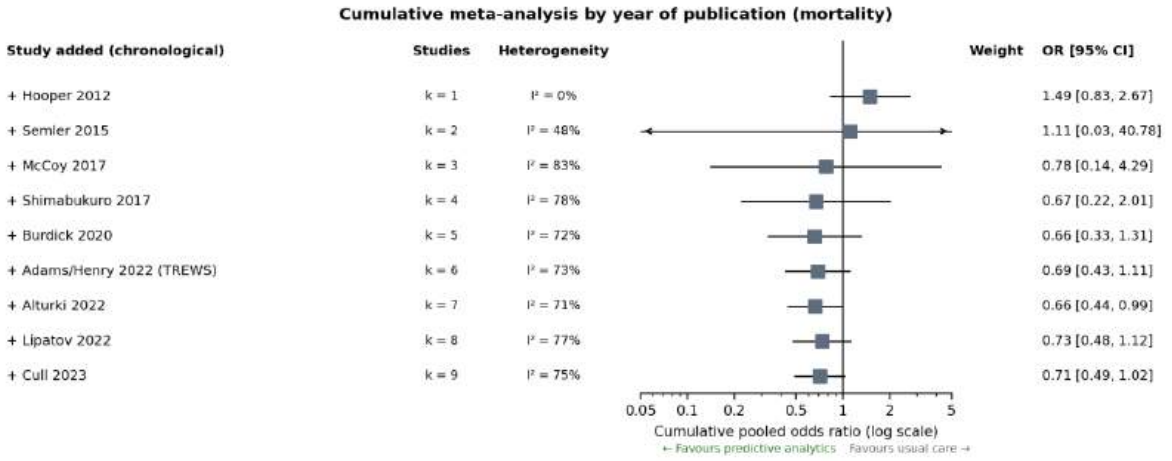


Figure 16. Cumulative random-effects meta-analysis of mortality ordered by year of publication.[43]

Table 7. Sensitivity analyses for the primary mortality outcome (odds ratio).

Analysis	k	OR (95% CI)	I^2 (%)	Interpretation
Primary analysis (DL + mKH)	9	0.71 (0.49–1.02)	75	Inconclusive
Common-effect (inverse variance)	9	0.72 (0.66–0.79)	75	Significant benefit
DerSimonian–Laird, Wald-type CI	9	0.71 (0.56–0.91)	75	Significant benefit
Cull 2023 adjusted OR instead of crude	9	0.71 (0.49–1.02)	75	Inconclusive
Excluding Semler 2015 (indirect intervention)	8	0.70 (0.46–1.06)	78	Inconclusive
Excluding TREWS (exposure = clinician confirmation)	8	0.70 (0.45–1.09)	75	Inconclusive
Excluding derived event counts (Hooper,	7	0.67 (0.43–1.05)	72	Inconclusive

Burdick)				
Excluding high/serious risk of bias	4	1.01 (0.44–2.32)	78	Inconclusive
Adult populations only (excluding Alturki 2022)	8	0.74 (0.49–1.11)	76	Inconclusive

Primary analysis: DerSimonian–Laird random effects with modified Hartung–Knapp CI. "Excluding high/serious risk of bias" retains Hooper 2012, Semler 2015, Lipatov 2022 and Cull 2023.

Meta-regression and reporting bias

Exploratory meta-regression did not identify publication year (slope -0.024 log-OR units per year, P = 0.50) or baseline control-group mortality (slope -0.0050 per percentage point, P = 0.71) as significant effect modifiers, and the L'Abbé plot showed that most studies lay below the line of equality across a wide range of baseline risks (3.9% to 29.9%). The contour-enhanced funnel plot showed no material asymmetry (Egger intercept -0.29, SE 1.21, P = 0.82); trim-and-fill imputed 1 study on the side of harm, yielding an adjusted OR of 0.73 (95% CI 0.52–1.04).[47,48]

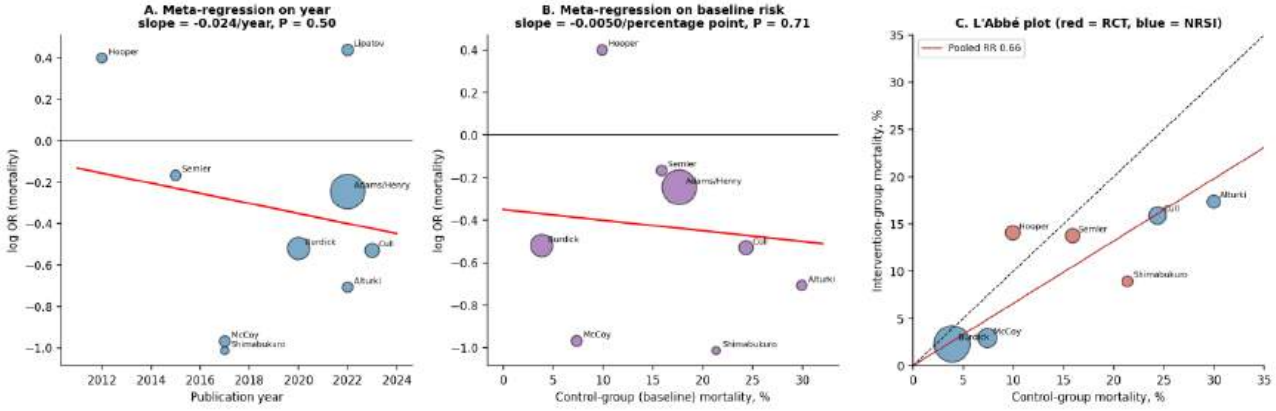


Figure 17. Exploratory random-effects meta-regression of log odds ratio for mortality on (A) year of publication and (B) baseline control-group mortality, and (C) L'Abbé plot of event rates (bubble size proportional to study precision or sample size).[46]

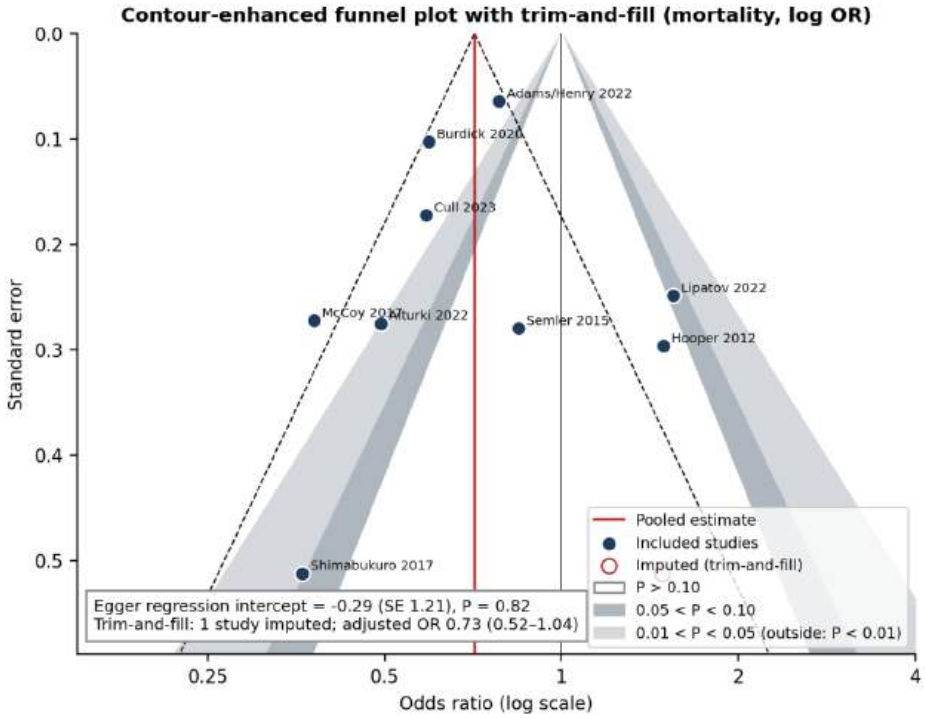


Figure 18. Contour-enhanced funnel plot of log odds ratios for mortality with trim-and-fill (open circle, imputed study). Shaded regions denote conventional significance contours around the null.[47,48]

Primary outcome: time to intervention

Five studies reported time to first antibiotic administration or order. Predictive analytics shortened the time to antibiotics by a mean of 1.83 hours, but with very high heterogeneity and an mKH interval including the null (MD -1.83 (95% CI -4.57 to 0.90); P = 0.136; I^2 = 88%; Wald CI -3.12 to -0.55, P = 0.005. The largest and most precise reductions were observed for TREWS (1.85 h in median time to first antibiotic order among confirmed alerts) and for the Epic Sepsis Model (median 90 versus 150 minutes), whereas the rule-based SIRS alert RCT showed no difference (6.0 versus 6.1 h). Results were unchanged after excluding the small surgical ICU study with critical risk of bias (MD -1.79 (95% CI -4.40 to 0.82)) or TREWS (MD -1.88 (95% CI -8.24 to 4.47)). The standardised analysis gave a small-to-moderate, non-significant effect (Hedges' g -0.30 (95% CI -0.90 to 0.29); I^2 = 90%. Two studies not poolable because of missing denominators reported concordant improvements: median time to antibiotics fell from 7.4 to 4.2 h after a surgical ICU best-practice advisory, and from 2.8 to 1.9 h after ICU implementation of the trajectory model.[23,55,60–63,66]

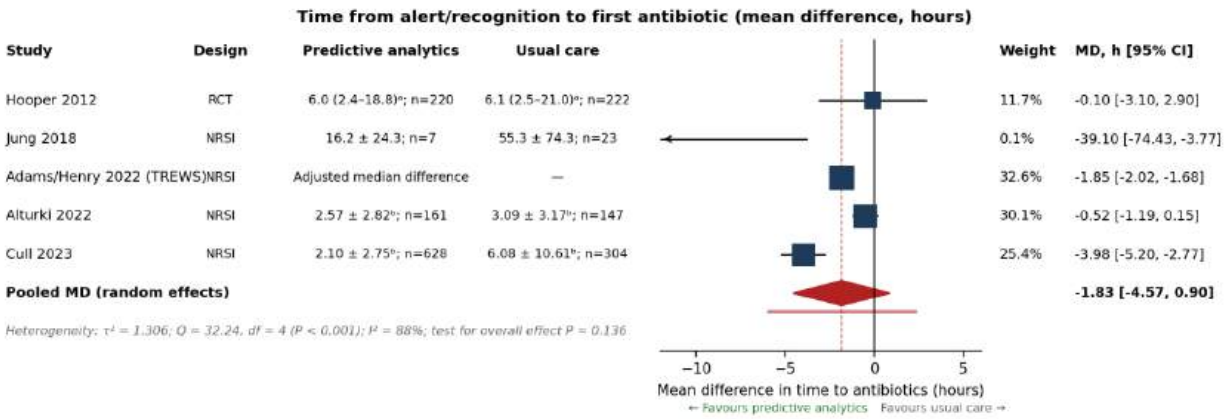


Figure 19. Forest plot of time to first antibiotic (mean difference, hours). [a] Median (interquartile range) converted to mean and SD; [b] mean ± SD estimated from median and interquartile range (Wan/Luo methods). Jung 2018 values are mean ± SD derived from reported SEM.[39,40]

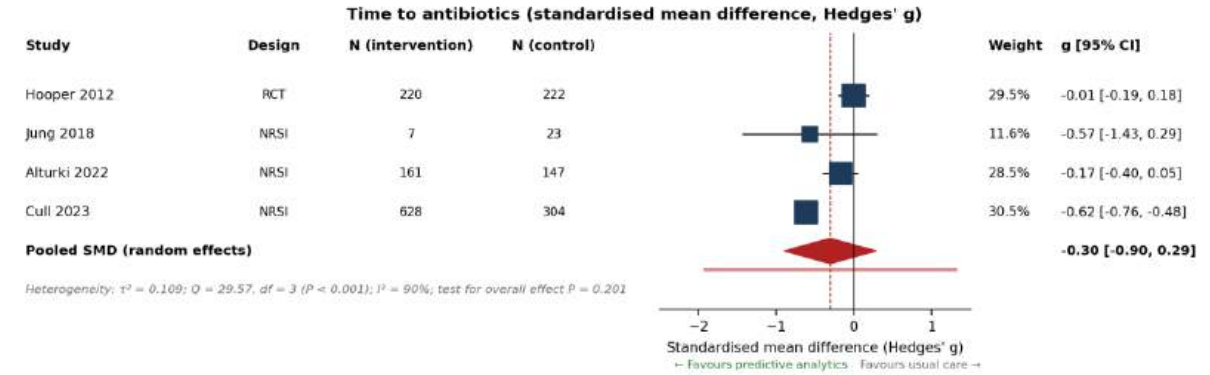


Figure 20. Forest plot of time to first antibiotic expressed as standardised mean difference (Hedges' g).[39]

Secondary outcomes: length of stay and guideline-concordant care

Across five studies, ICU LOS was shorter by 1.10 days with predictive analytics, without reaching significance (MD -1.10 (95% CI -2.92 to 0.72); I^2 = 91%; PI -5.68 to 3.48); the reductions were largest in the ML trial (−2.09 days) and the trajectory-model implementation (−1.8 days). Hospital LOS did not differ (MD 0.27 (95% CI -3.61 to 4.14); k = 3; I^2 = 68%); two further ML implementation studies reported shorter crude hospital LOS (3.27 versus 4.83 days and 2.92–3.19 versus 3.35 days) without dispersion measures.[55,57–61,66]

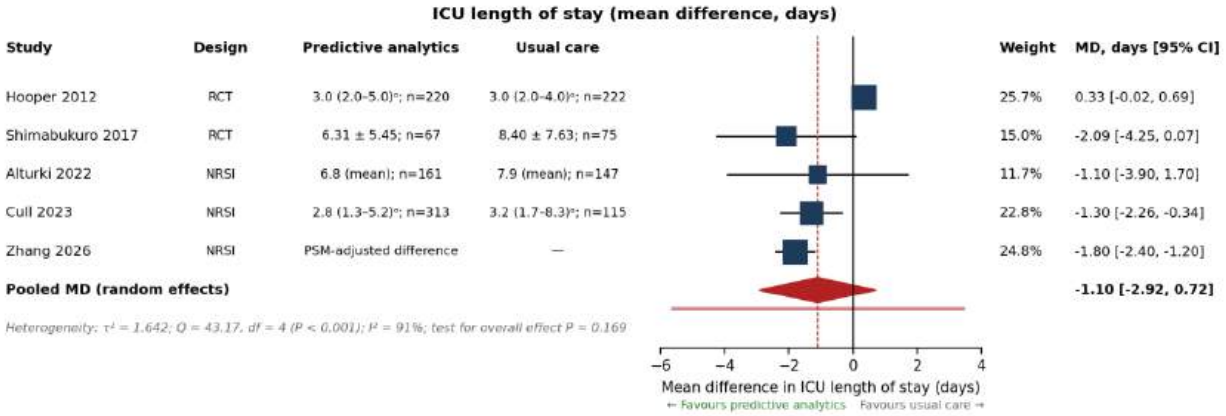


Figure 21. Forest plot of ICU length of stay (mean difference, days).[39]

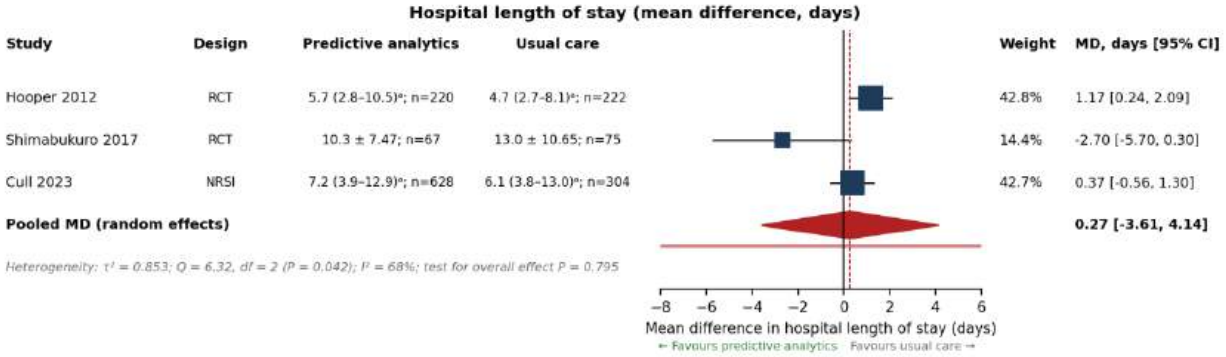


Figure 22. Forest plot of hospital length of stay (mean difference, days).[39]

Three non-randomised studies reported adherence to guideline-concordant care. Predictive analytics increased the odds of timely recommended interventions or order-set use (OR 1.43 (95% CI 0.85–2.41); Wald CI 1.13–1.81; I^2 = 1%; Figure 21), with low heterogeneity. In the RCT of an electronic evaluation and management tool, time to completion of the 6-hour bundle did not differ (HR 1.98, 95% CI 0.75–5.20), plausibly because the tool was used to place orders in only 28% of eligible cases.[56,60,64,65]

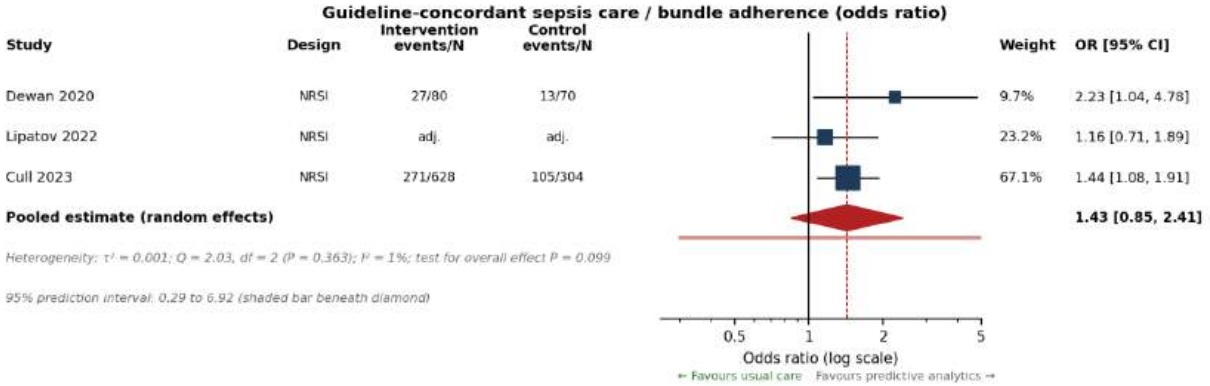


Figure 23. Forest plot of guideline-concordant sepsis care (timely recommended interventions, order-set use or bundle compliance; odds ratio). Values above 1 favour predictive analytics.[64]

Diagnostic performance of ICU sepsis prediction algorithms

Fifteen validation datasets contributed to the pooled discrimination analysis. The summary AUROC was 0.864 (95% CI 0.816–0.901), with extreme heterogeneity (I^2 = 99.7%, τ^2 [logit] = 0.414) and a 95% PI of 0.60–0.96. Discrimination did not differ between ICU-only (0.854, 0.790–0.900) and mixed settings (0.879, 0.710–0.956; P = 0.64) or between internal (0.874) and external validation (0.858; P = 0.59). Excluding the proprietary Epic models increased the pooled AUROC to 0.877 (0.827–0.914; k = 12). The lowest values were observed for the first-generation Epic Sepsis Model (0.63) and for international external validation of a deep-learning model (0.761), whereas the highest were reported for gradient-boosted models evaluated on vital-sign data (0.92–0.948). The funnel plot suggested possible small-study effects (Egger P = 0.10).[17–22,58,60,67–73]

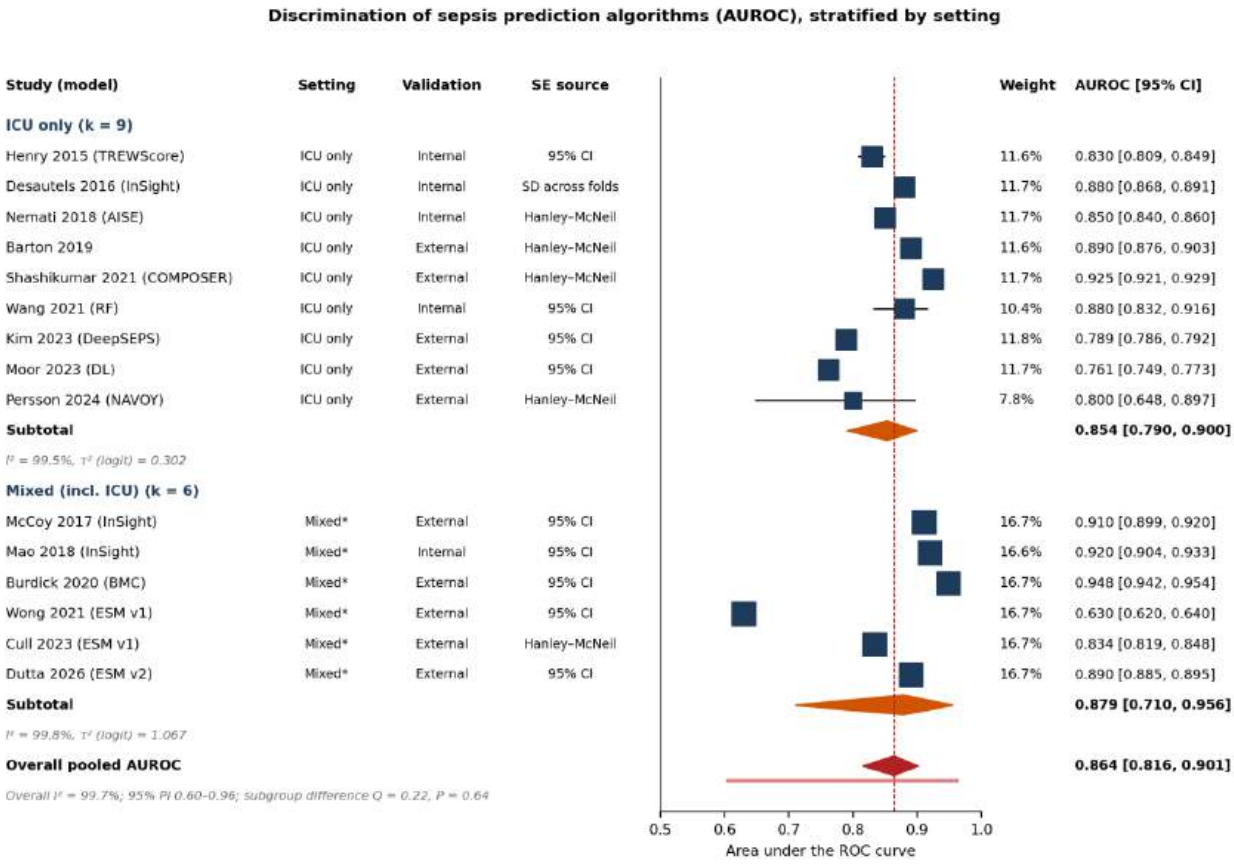


Figure 24. Forest plot of discrimination (AUROC) of sepsis prediction algorithms stratified by setting (ICU-only versus mixed settings including ICU; * mixed). Pooling was performed on the logit scale with random effects.[44,45]

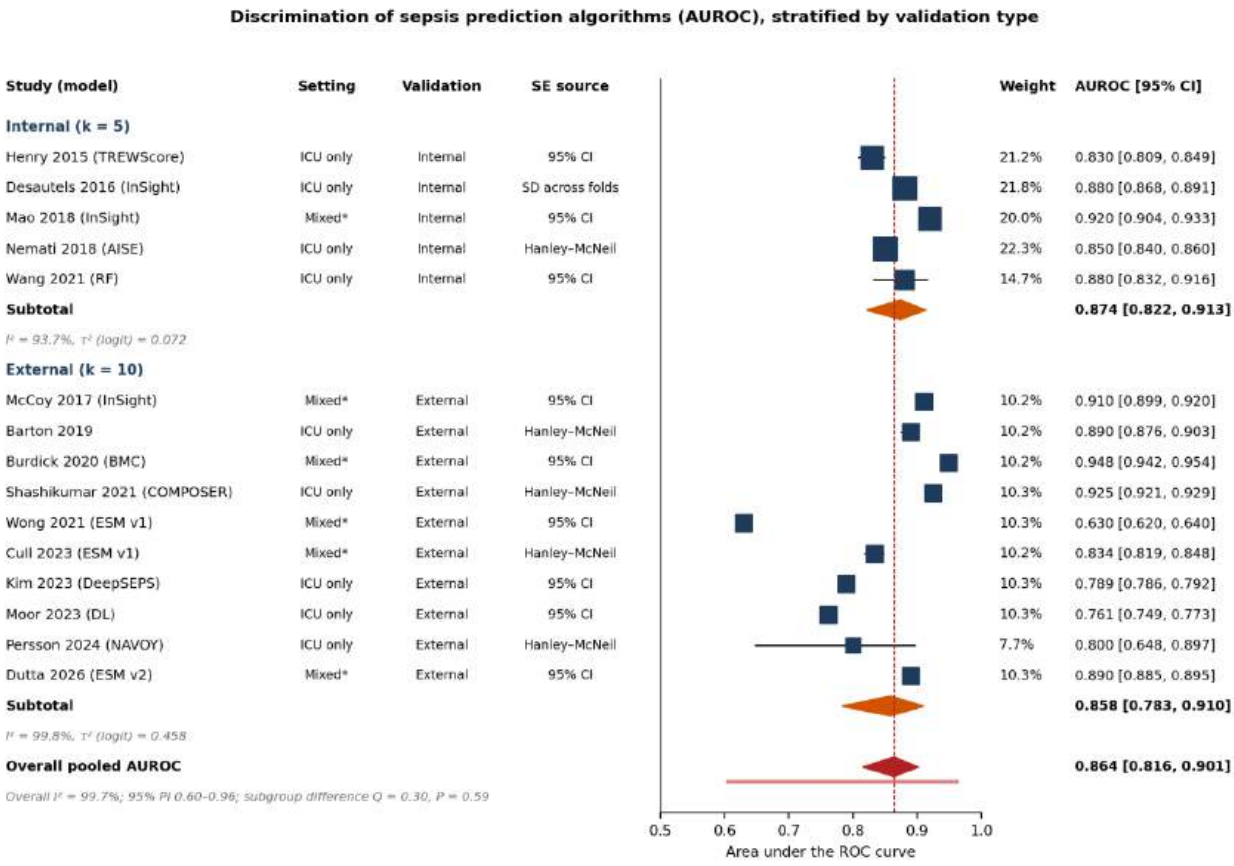


Figure 25 Forest plot of discrimination (AUROC) stratified by validation type (internal versus external, temporal or prospective).[44]

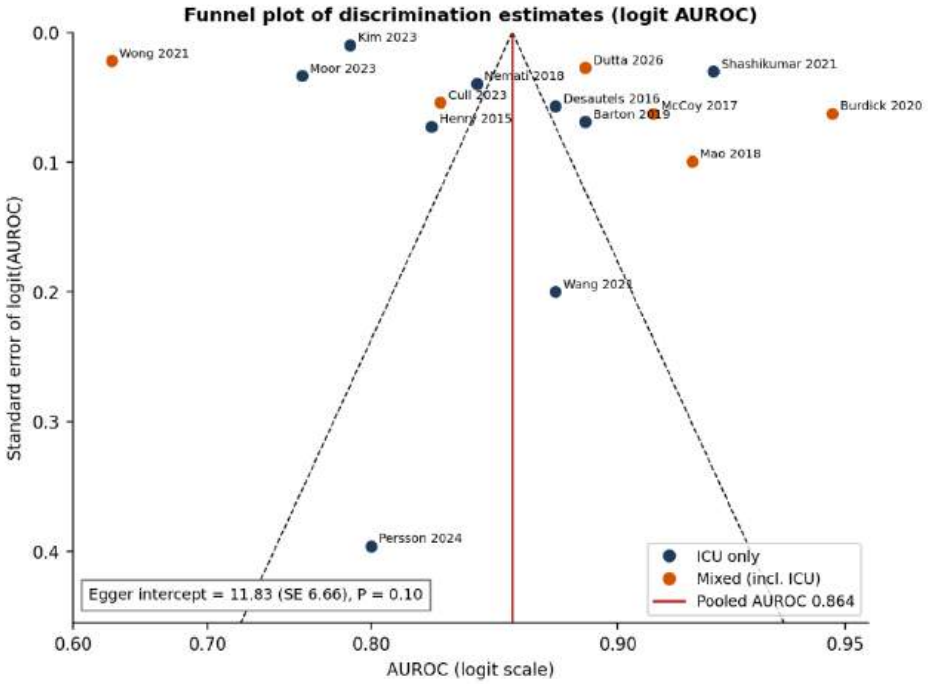


Figure 26 Funnel plot of logit-transformed AUROC estimates by setting with Egger's regression test.[47]

At the implemented alert thresholds, sensitivity varied widely (0.33 for the first-generation Epic Sepsis Model at a score of 6 or higher, 0.80 for NAVOY and 0.86 for the Epic Sepsis Model at

a locally optimised threshold), precluding meaningful pooling ($I^2 > 99\%$), whereas specificity was more consistent (pooled 0.82, 95% CI 0.77–0.86). Positive predictive values were generally low (11–46%), implying between one and eight false alerts per true alert. Reported lead times ranged from 3 to 28 hours before onset, and median lead time before the first antibiotic order was 12.2 hours for COMPOSER in the ICU.[17,20–22,60,71,73]

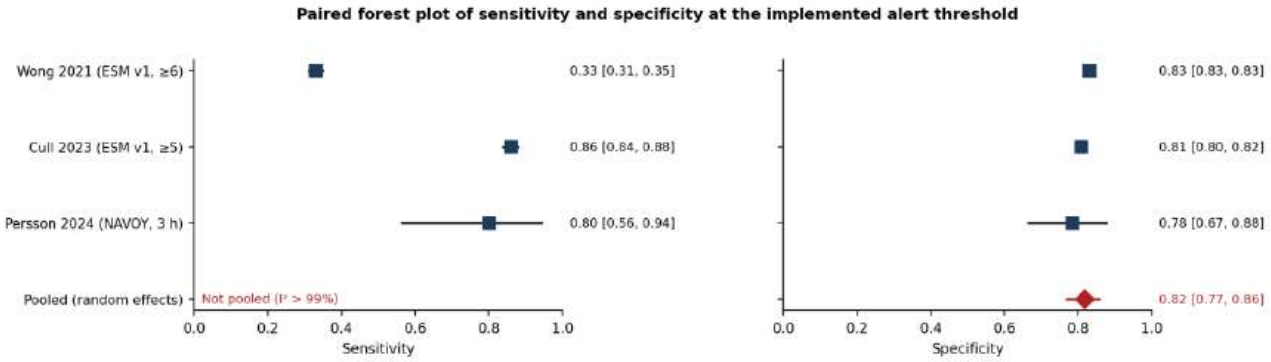


Figure 27 Paired forest plot of sensitivity and specificity (exact 95% CI) at the implemented alert threshold for studies with derivable 2 × 2 data.[21,22,60]

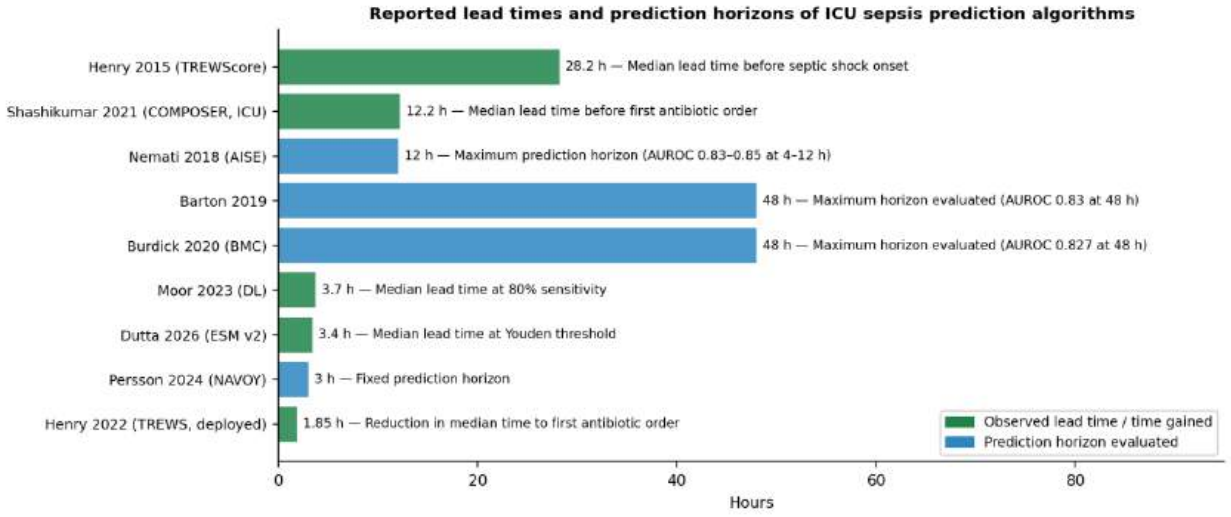


Figure 28 Reported lead times and evaluated prediction horizons of ICU sepsis prediction algorithms.[17,19–21,23,68,69,71,73]

Table 8. Operating characteristics and alert burden at validation or deployment thresholds.

Study	Algorithm	Threshold / unit	Sens.	Spec.	PPV	NPV	Alert burden / lead time
Hooper 2012	Rule-based SIRS	Modified SIRS; MICU patients	0.99	0.82	0.41	0.97	Applied to all MICU admissions
Lipatov 2022	Rule-based sniffer	Alert; MICU/ED	0.80	0.77	0.28	0.97	3,424 unique alerts; 1,131 confirmed sepsis

Dewan 2020	Rule-based CDS (PICU)	Sepsis episodes	0.93	0.96	0.46	>0.99	45.0 sepsis events/1,000 patient-days
Mao 2018	InSight (gradient boosting)	Fixed sensitivity 0.80	0.80	0.95	NR	NR	Detection (0 h)
McCoy 2017	InSight	Sepsis-3	0.83	0.96	NR	NR	Retrospective local evaluation
Shashikumar 2021	COMPOSER	External ICU; patient/window level	0.91	0.91	0.24	0.99	Median 12.2 h before antibiotics; ~20% indeterminate
Wong 2021	ESM v1	Score ≥ 6; hospitalisation	0.33	0.83	0.12	0.95	Alerts in 18% of hospitalisations; 67% of sepsis missed
Cull 2023	ESM v1	Score ≥ 5; encounter	0.86	0.81	0.34	0.98	Cut-point derived locally
Moor 2023	Deep learning	80% sensitivity	0.80	NR	NR	NR	1.4 false alerts per true alert; 3.7 h lead
Persson 2024	NAVOY Sepsis	3 h ahead; patient	0.80	0.78	NR	NR	Accuracy 0.79 (0.70–0.88)
Dutta 2026	ESM v2	Youden threshold 9; Sepsis-3	NR	NR	0.11	NR	Median lead time 3.4 h

Values as reported or derived from reported counts (Wong 2021: 843/2,552 sepsis identified; 6,971 alerted of 38,455). NR, not reported; NPV, negative predictive value; PPV, positive predictive value.

Summary of meta-analyses

All quantitative syntheses are summarised in Table. With the exception of the count-based RR and the absolute RD for mortality and the ML subgroup, no pooled clinical outcome reached significance under the conservative mKH method, although point estimates consistently favoured predictive analytics for mortality, time to antibiotics, ICU LOS and guideline-concordant care.[36,43]

Table 9. Summary of pooled estimates for all outcomes.

Outcome (effect measure)	k (design)	Pooled estimate (mKH 95% CI)	Wald 95% CI	I^2 / τ^2	95% PI	P (mKH)
Mortality (OR)	9 (3 RCT, 6 NRSI)	0.71 (0.49–1.02)	0.56–0.91	75% / 0.083	0.34–1.49	0.063
Mortality (RR, counts)	7 (3 RCT, 4 NRSI)	0.66 (0.46–0.94)	0.52–0.84	62% / 0.059	0.33–1.33	0.029
Mortality (RD, % points)	9 (3 RCT, 6 NRSI)	-3.8 (-6.8 to -0.8)	-5.8 to -1.8	72% / 0.00048	-9.5 to 1.9	0.019
Mortality, ML/predictive (OR)	5 (1 RCT, 4 NRSI)	0.60 (0.42–0.84)	0.47–0.76	69% / 0.043	0.28–1.29	0.015
Mortality, rule-based (OR)	4 (2 RCT, 2 NRSI)	0.99 (0.42–2.37)	0.58–1.70	75% / 0.223	0.09–10.37	0.979
Time to antibiotics (MD, h)	5 (1 RCT, 4 NRSI)	-1.83 (-4.57 to 0.90)	-3.12 to -0.55	88% / 1.306	-6.03 to 2.36	0.136
Time to antibiotics (SMD)	4 (1 RCT, 3 NRSI)	-0.30 (-0.90 to 0.29)	-0.67 to 0.06	90% / 0.109	-1.94 to 1.33	0.201
ICU LOS (MD, days)	5 (2 RCT, 3 NRSI)	-1.10 (-2.92 to 0.72)	-2.39 to 0.19	91% / 1.642	-5.68 to 3.48	0.169
Hospital LOS (MD, days)	3 (2 RCT, 1 NRSI)	0.27 (-3.61 to 4.14)	-1.06 to 1.60	68% / 0.853	-14.3 to 14.8	0.795
Guideline-concordant care (OR)	3 (NRSI)	1.43 (0.85–2.41)	1.13–1.81	1% / 0.001	0.29–6.92	0.099
Discrimination (AUROC)	15 validation datasets	0.864 (0.816–0.901)	0.820–0.899	99.7% / 0.414 (logit)	0.60–0.96	< 0.001

mKH, modified Hartung–Knapp–Sidik–Jonkman; NRSI, non-randomised study of interventions; PI, prediction interval (t distribution, k − 2 degrees of freedom).

Table 10. GRADE summary of findings: predictive analytics for early sepsis detection compared with usual care in ICU settings.[46,49,50]

Outcome	Studies (participants)	Relative effect (95% CI)	Anticipated absolute effect (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Certainty	Comment
All-cause mortality	9 (3 RCT, 6 NRSI); 30,144	OR 0.71 (0.49–1.02)	159 per 1,000 with usual care; 40 fewer per 1,000 (73 fewer to 3 more)	Serious [a]	Serious[b]	Not serious	Serious[c]	Undetected	Very low ⨁◯◯◯	Effect disappears when high/serious RoB studies are excluded
Mortality — ML/predictive systems	5 (1 RCT, 4 NRSI); 27,037	OR 0.60 (0.42–0.84)	58 fewer per 1,000 (85 fewer to 22 fewer)	Serious	Serious	Not serious	Not serious	Undetected	Very low ⨁◯◯◯	Subgroup credibility low (P interaction 0.089)
Mortality — RCTs only	3 RCT; 991	OR 0.85 (0.16–4.43)	21 fewer per 1,000 (130 fewer to 297 more)	Serious	Serious	Serious[d]	Very serious	Undetected	Very low ⨁◯◯◯	Two of three trials tested rule-based tools

Outcome	Studies (participants)	Relative effect (95% CI)	Anticipated absolute effect (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Certainty	Comment
Time to first antibiotic	5 (1 RCT, 4 NRSI); 1,712 + TREWS cohort	—	MD 1.83 h shorter (4.57 shorter to 0.90 longer)	Serious	Very serious	Serious[e]	Serious	Undetected	Very low ⨁◯◯◯	Different time-zero definitions
ICU length of stay	5 (2 RCT, 3 NRSI); 1,320 + implementation cohort	—	MD 1.10 days shorter (2.92 shorter to 0.72 longer)	Serious	Very serious	Not serious	Serious	Undetected	Very low ⨁◯◯◯	—

Outcome	Studies (participants)	Relative effect (95% CI)	Anticipated absolute effect (95% CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Certainty	Comment
Hospital length of stay	3 (2 RCT, 1 NRSI); 1,516	—	MD 0.27 days longer (3.61 shorter to 4.14 longer)	Serious	Serious	Not serious	Very serious	Undetected	Very low ⨁◯◯◯	—
Guideline-concordant care	3 NRSI; 3,032	OR 1.43 (0.85–2.41)	265 per 1,000 with usual care; 75 more per 1,000 (31 fewer to 200 more)	Serious	Not serious	Serious[e]	Serious	Undetected	Very low ⨁◯◯◯	Heterogeneous adherence definitions
Discrimination (AUROC)	15 validation datasets (>860,000 patients/encounters)	AUROC 0.864 (0.816–0.901)	95% PI 0.60–0.96	Serious	Serious[b]	Not serious	Not serious	Suspected (Egger P = 0.10)	Low ⨁⨁◯◯	Cohort accuracy studies start at high certainty

[a] Most studies at serious risk of bias (confounding, unblinded alerts). [b] Substantial unexplained heterogeneity ($I^2 > 60\%$) and wide prediction interval. [c] The 95% CI (modified Hartung–Knapp) includes no effect. [d] Rule-based interventions evaluated in two trials may not represent contemporary ML systems. [e] Surrogate process outcome with heterogeneous definitions. Baseline risks represent the median control-group risk of included studies. Certainty: ⨁⨁⨁⨁ high; ⨁⨁⨁◯ moderate; ⨁⨁◯◯ low; ⨁◯◯◯ very low.

DISCUSSION

Principal findings

This systematic review and meta-analysis of 26 studies (4 RCTs, 10 comparative implementation studies and 12 diagnostic validation cohorts) found that predictive analytics for early sepsis detection in ICU and mixed hospital settings including ICU patients were associated with a 29% reduction in the odds of death (OR 0.71) that did not reach significance under conservative inference, but which was supported by significant reductions in the count-based risk ratio and the absolute risk difference (approximately 3.8 percentage points). The signal was concentrated in ML-based or predictive systems (OR 0.60; 58 fewer deaths per 1,000) and was absent for rule-based alerts. Predictive analytics shortened time to antibiotics by about 1.8 hours and ICU LOS by about 1.1 days, with wide uncertainty, and pooled discrimination was good (AUROC 0.864) yet highly heterogeneous. Because the benefit disappeared when studies at serious risk of bias were excluded and all outcomes were rated as very low to low certainty, the hypothesis that predictive analytics reduces mortality (H1) is supported only weakly, H2 is directionally but not statistically supported, H3 is suggested, and H4 is supported.[30,55–60]

Comparison with previous systematic reviews

Our results extend, and partly reconcile, previous syntheses. The Bayesian meta-analysis of automated alerting reported a beneficial effect overall (RR 0.71) and a larger effect for ML-based methods (RR 0.56), but a non-significant effect in the ICU (RR 0.90); our ICU-centred analysis, which adds prospective and multicentre studies published after 2021, found a similar overall magnitude with a comparable ML advantage, while the ICU-only subgroup (OR 0.73) remained imprecise. Emergency department reviews reported a mortality RR of 0.81 for sepsis alert systems, and a review of digital alerting found reductions in hospital and ICU LOS without a significant mortality effect, consistent with our LOS findings. The Cochrane review restricted to three ICU RCTs could not pool data and graded all outcomes as very low certainty, a conclusion that our larger evidence base does not overturn.[27–29,74]

For discrimination, earlier diagnostic reviews reported wide AUROC ranges (0.68–0.99 in ICUs) and precluded pooling because of heterogeneous sepsis definitions and prediction windows; a

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

2025 methodological review demonstrated that median AUROC declined from 0.886 at 6 hours to 0.783 under full-window external validation, and a 2026 meta-analysis of externally validated Epic tools reported a pooled AUROC of only 0.65 for the Epic Sepsis Model with I^2 of at least 93%. Our pooled AUROC of 0.864, falling to a lower prediction bound of 0.60, mirrors this gap between development and real-world performance, and the improvement in the second-generation, locally trained Epic model (AUROC 0.89) illustrates the value of local recalibration.[24–26,73,75,76]

Table 11. Comparison with previous systematic reviews and meta-analyses (not included in the quantitative synthesis).

Review	Scope / setting	Included studies	Key pooled finding	Relation to present review
Fleuren 2020	ML sepsis prediction, any hospital setting	28 (130 models)	AUROC 0.68–0.99 in ICU; only 3 clinical implementations	Present review adds 2020–2026 external and prospective validations
Moor 2021	ML sepsis onset prediction, adult ICU	22	Meta-analysis precluded by heterogeneity	Present review pools 15 datasets on logit scale
Joshi 2019	Digital alerting, adult inpatients (ICU-only excluded)	16	Hospital LOS −1.31 days; ICU LOS −0.77 days; mortality NS	Concordant LOS direction; present review focuses on ICU
Warttig 2018 (Cochrane)	Automated monitoring vs standard care, ICU RCTs	3 RCTs	No pooling; very low certainty	Adds RCT of ML and non-randomised evidence
Bao 2022	Automated alerts vs usual care, all	36 (6 RCTs)	RR 0.71 overall; ICU RR 0.90 (NS); ML RR	Similar overall and ML effects; ICU

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

	settings		0.56	subgroup still imprecise
Kim 2024	Sepsis alerts, adult ED	22	Mortality RR 0.81; time to antibiotics SMD −0.34	ED evidence excluded here; effect sizes comparable
Wang 2025	Methodology of real-time sepsis models	91	AUROC 0.886 (6 h) → 0.783 full-window external	Explains heterogeneity and optimism in AUROC
Abbas 2025	AI models, adult inpatients	52	AUC 0.79–0.96; limited prospective validation	Consistent discrimination range
Patel 2026	Externally validated Epic CDS tools	22	ESM pooled AUROC 0.65; $I^2 \geq 93\%$	Supports proprietary-model underperformance
Mazza 2026	AI mortality/outcome prediction in ICU sepsis	75	Median ΔAUROC +0.108 vs severity scores; 45 high RoB	Prognostic (not detection) focus; similar RoB concerns
Zhang 2026	AI sepsis prediction, ED	36 (98 models)	Pooled AUROC 0.87	Comparable to ICU estimate (0.864)
Driver 2026	EMR sepsis alerts, paediatrics	12	Improved time to antibiotics; mortality uncertain	Consistent with paediatric ICU studies included here

ED, emergency department; NS, not significant; RoB, risk of bias.

From discrimination to outcomes: the alert-to-action pathway

The modest and uncertain translation of good discrimination into improved outcomes is best understood through the alert-to-action pathway. The biological rationale is strong: observational data consistently show that each hour of delay in antimicrobial therapy increases mortality,

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

particularly in septic shock, and MOSAICS II identified delayed antibiotics as the only bundle element independently associated with death in Asian ICUs. Predictive analytics can only realise this benefit if alerts arrive before clinical recognition and prompt different clinical behaviour.[6,10,11,13,14]

Several included studies illustrate where the chain breaks. In the TREWS programme, 89% of alerts were evaluated but only 38% of evaluated alerts were confirmed, and the mortality benefit was confined to patients whose alerts were confirmed within 3 hours, with a 1.85-hour reduction in time to the first antibiotic order. In the RCT of an electronic sepsis evaluation tool, providers used the tool to enter orders in only 28% of cases, and in the NAVOY randomised validation very few patients had fluids, cultures or antibiotics initiated at the time of the alert. Conversely, when predictions are generated after clinicians have already recognised sepsis, discrimination becomes illusory: excluding post-recognition predictions reduced the AUROC of the Epic Sepsis Model from 0.62 to 0.47. These observations indicate that adoption, timing and workflow integration determine clinical value at least as much as statistical performance.[21,23,30,31,56]

Why machine-learning systems may outperform rule-based alerts

The divergence between ML-based and rule-based systems is clinically plausible. Rule-based alerts typically rely on SIRS criteria, which are present in many non-infected critically ill patients and discriminate poorly in the ICU; consequently, rule-based surveillance has low positive predictive value (14.6% in an emergency department implementation and 28% in a medical ICU), generating alert fatigue that undermines response. ML models integrate trajectories of vital signs and laboratory values, can learn patient-specific baselines and, in the case of conformal approaches such as COMPOSER, abstain from predictions in unfamiliar data, reducing false alarms. Nevertheless, ML is not uniformly superior: the first-generation Epic Sepsis Model alerted on 18% of hospitalisations while missing 67% of sepsis cases, and even contemporary models have positive predictive values of only 11–26%, creating a substantial number needed to evaluate.[8,20,22,54,65,73,77]

Sources of heterogeneity

Heterogeneity was substantial for clinical outcomes and extreme for discrimination. Plausible sources include differing sepsis definitions (Sepsis-3, SIRS-based severe sepsis,

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

surveillance or diagnostic-code definitions), time-zero conventions for antibiotic timing (alert, recognition, triage or arrival), prediction horizons (0–48 hours), sepsis prevalence (3–23%), case-mix (adult versus paediatric, medical versus surgical ICUs), intensity of the accompanying response protocol and the evaluation unit (patient, encounter or hourly window). Very large validation cohorts produce extremely precise estimates, so that I^2 approaches 100% even when absolute differences in AUROC are modest; the prediction interval is therefore more informative than I^2 for decision-makers.[7,26,44,45]

Implications for clinical practice

For ICUs considering predictive analytics, our findings support four practical principles. First, prefer ML-based systems with demonstrated external or prospective validation, and perform local silent validation and recalibration before activating alerts, as exemplified by the prospective silent deployment of the second-generation Epic model. Second, couple alerts to an explicit response protocol (bedside evaluation within a defined time, culture and lactate acquisition, antimicrobial decision) and monitor adoption metrics such as the proportion of alerts evaluated and confirmed. Third, track alert burden, positive predictive value and lead time before clinical recognition rather than relying on AUROC alone. Fourth, embed the programme within the performance-improvement framework recommended by the Surviving Sepsis Campaign.[15,16,23,31,73]

These principles are particularly relevant for health systems in low- and middle-income countries, including Indonesia, where sepsis bundle compliance is low, intensivist coverage is limited and EHR adoption is expanding. Because no included study originated from South-East Asia, models developed in high-income settings should not be assumed to transport; rule-based screening adapted to local workflows may be a pragmatic first step in primary care-to-hospital referral pathways, whereas ML systems should be introduced only after local validation reported according to TRIPOD+AI and appraised with PROBAST+AI.[1,6,37,42]

Implications for research

The evidence base remains dominated by single-centre, open-label or uncontrolled designs. Pragmatic, multicentre randomised designs embedded in the EHR — such as the factorial SEPTIC trials comparing nurse-facing and prescriber-facing alerts and the PRECISE precision-resuscitation

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

trial operationalising an ML algorithm across six hospitals and seventeen ICUs — offer a feasible template. Future studies should report the full alert-to-action cascade, alert burden, calibration, decision-curve analysis and outcomes stratified by sex, age and ethnicity, and should adhere to TRIPOD+AI and CONSORT-AI standards.[37,38,78,79]

Strengths

Strengths of this review include its explicit ICU focus, the restriction to evidence from the last 15 years, the joint synthesis of process, patient and accuracy outcomes, the use of the modified Hartung–Knapp method and prediction intervals appropriate for few heterogeneous studies, design-specific risk-of-bias tools, extensive sensitivity analyses and transparent documentation of all data transformations.[32,36,43]

CONCLUSIONS

Predictive analytics algorithms for early sepsis detection in the ICU, particularly ML-based early warning systems, discriminate sepsis well and show a consistent, clinically meaningful signal towards earlier antimicrobial therapy, shorter ICU stay and lower mortality. However, the certainty of evidence is low to very low, the apparent mortality benefit depends largely on non-randomised studies at serious risk of bias, and performance varies widely across institutions. Predictive analytics should therefore be implemented as part of structured sepsis performance-improvement programmes with local validation, explicit response protocols and continuous monitoring of alert burden and adoption, while adequately powered pragmatic multicentre RCTs — including in low- and middle-income ICUs — are urgently needed to establish effect on patient-centred outcomes.[15,16,30,57]

DECLARATIONS

Author contributions (CRediT). Yudha Viantoro: conceptualisation, methodology, literature search, data curation, formal analysis, visualisation and writing — original draft. Fifa Yuniarmi: validation, data extraction verification, risk-of-bias assessment and writing — review and editing. Both authors approved the final manuscript and agree to be accountable for all aspects of the work.

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

Funding. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests. The authors declare no competing interests.

Data availability. The extraction dataset, derivation log (Supplementary Table S2) and analysis code are available from the corresponding author on reasonable request.

Ethics approval and consent to participate. Not applicable; this study synthesised published, de-identified aggregate data.

Acknowledgements. None.

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

REFERENCES

1. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. Lancet. 2020;395(10219):200-11. doi:10.1016/s0140-6736(19)32989-7. Available from: https://consensus.app/papers/details/0c127ae4bda65858b4b9aa980e2f4bb1/?utm_source
2. Gray AP, Chung E, Hsu RL, Araki DT, Gershberg Hayoon A, Davis Weaver N, et al. Global, regional, and national sepsis incidence and mortality, 1990–2021: a systematic analysis. Lancet Glob Health. 2025;13(12):e2013-e2026. doi:10.1016/s2214-109x(25)00356-0. Available from: https://consensus.app/papers/details/e98bea434f7e52a3b158323fb8928e0f/?utm_source
3. Fleischmann-Struzek C, Mellhammar L, Rose N, Cassini A, Rudd KE, Schlattmann P, et al. Incidence and mortality of hospital- and ICU-treated sepsis: results from an updated and expanded systematic review and meta-analysis. Intensive Care Med. 2020;46(8):1552-62. doi:10.1007/s00134-020-06151-x. Available from: https://consensus.app/papers/details/0d75b64271ef5ca093f08d9270c91ed7/?utm_source
4. Sakr Y, Jaschinski U, Wittebole X, Szakmany T, Lipman J, Ñamendys-Silva SA, et al. Sepsis in Intensive Care Unit Patients: Worldwide Data From the Intensive Care over Nations Audit. Open Forum Infect Dis. 2018;5(12):ofy313. doi:10.1093/ofid/ofy313. Available from: https://consensus.app/papers/details/70d73fb72ed651fea61cd319e58e395f/?utm_source
5. Markwart R, Saito H, Harder T, Tomczyk S, Cassini A, Fleischmann-Struzek C, et al. Epidemiology and burden of sepsis acquired in hospitals and intensive care units: a systematic review and meta-analysis. Intensive Care Med. 2020;46(8):1536-51. doi:10.1007/s00134-020-06106-2. Available from: https://consensus.app/papers/details/25fab809f8eb5febab2c1f58897b75bb/?utm_source
6. Li A, Ling L, Qin H, Arabi YM, Myatra SN, Egi M, et al. Epidemiology, Management, and Outcomes of Sepsis in ICUs among Countries of Differing National Wealth across Asia. Am J Respir Crit Care Med. 2022;206(9):1107-16. doi:10.1164/rccm.202112-2743oc. Available from: https://consensus.app/papers/details/1f34b35d63145551aa0945774db47c08/?utm_source

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

7. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801-10. doi:10.1001/jama.2016.0287. Available from: https://consensus.app/papers/details/a3e5dbdcba4358048383e4f7c20c9d77/?utm_source
8. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of Clinical Criteria for Sepsis. JAMA. 2016;315(8):762-74. doi:10.1001/jama.2016.0288. Available from: https://consensus.app/papers/details/e5971199c0605572b2db041dcc382911/?utm_source
9. Schlapbach LJ, Watson RS, Sorce LR, Argent AC, Menon K, Hall MW, et al. International Consensus Criteria for Pediatric Sepsis and Septic Shock. JAMA. 2024;331(8):665-74. doi:10.1001/jama.2024.0179. Available from: https://consensus.app/papers/details/d8dd30b162d557b698d9f089ae58cae0/?utm_source
10. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. N Engl J Med. 2017;376(23):2235-44. doi:10.1056/nejmoa1703058. Available from: https://consensus.app/papers/details/2056794495f85b98a5a8e7fd7fe9666d/?utm_source
11. Liu VX, Fielding-Singh V, Greene JD, Baker JM, Iwashyna TJ, Bhattacharya J, et al. The Timing of Early Antibiotics and Hospital Mortality in Sepsis. Am J Respir Crit Care Med. 2017;196(7):856-63. doi:10.1164/rccm.201609-1848oc. Available from: https://consensus.app/papers/details/2591e8f43db45ffbb4fc87fa52fe03e8/?utm_source
12. Peltan ID, Brown SM, Bledsoe JR, Sorensen J, Samore MH, Allen TL, et al. ED Door-to-Antibiotic Time and Long-term Mortality in Sepsis. Chest. 2019;155(5):938-46. doi:10.1016/j.chest.2019.02.008. Available from: https://consensus.app/papers/details/60d0146ccfa953f68c33b270b406f807/?utm_source
13. Im Y, Kang D, Ko RE, Lee YJ, Lim SY, Park S, et al. Time-to-antibiotics and clinical outcomes in patients with sepsis and septic shock: a prospective nationwide multicenter cohort study. Crit Care. 2022;26(1):19. doi:10.1186/s13054-021-03883-0. Available from: https://consensus.app/papers/details/c7ba838985115b59920eb45b970d1093/?utm_source

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

14. Rüddel H, Thomas-Rüddel DO, Reinhart K, Bach F, Gerlach H, Lindner M, et al. Adverse effects of delayed antimicrobial treatment and surgical source control in adults with sepsis: results of a planned secondary analysis of a cluster-randomized controlled trial. Crit Care. 2022;26(1):51. doi:10.1186/s13054-022-03901-9. Available from: https://consensus.app/papers/details/6d0b851055ff5102bcb5324573cffcb9/?utm_source
15. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Crit Care Med. 2021;49(11):e1063-e1143. doi:10.1097/ccm.0000000000005337. Available from: https://consensus.app/papers/details/45bc29ed2f5653058510c321bbb012ea/?utm_source
16. Prescott HC, Antonelli M, Alhazzani W, Møller MH, Alshamsi F, Azevedo LCP, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2026. Crit Care Med. 2026. Published online ahead of print. doi:10.1097/ccm.0000000000007075. Available from: https://consensus.app/papers/details/c1906ee63bd55715ad1e4febcce86bcb/?utm_source
17. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med. 2015;7(299):299ra122. doi:10.1126/scitranslmed.aab3719. Available from: https://consensus.app/papers/details/7fd771cb47d458d5887fbe6b0436d831/?utm_source
18. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. JMIR Med Inform. 2016;4(3):e28. doi:10.2196/medinform.5909. Available from: https://consensus.app/papers/details/ddaa9e5a9f4a5d7483969df28c113cdf/?utm_source
19. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Crit Care Med. 2018;46(4):547-53. doi:10.1097/ccm.0000000000002936. Available from: https://consensus.app/papers/details/149da1e6158c57b3a2bb9551c40469ee/?utm_source
20. Shashikumar SP, Wardi G, Malhotra A, Nemati S. Artificial intelligence sepsis prediction algorithm learns to say “I don’t know”. NPJ Digit Med. 2021;4(1):134. doi:10.1038/s41746-

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

021-00504-6. Available from: https://consensus.app/papers/details/4560689e85765b94840a3086c881b061/?utm_source

21. Persson I, Macura A, Becedas D, Sjövall F. Early prediction of sepsis in intensive care patients using the machine learning algorithm NAVOY® Sepsis, a prospective randomized clinical validation study. J Crit Care. 2024;80:154400. doi:10.1016/j.jcrc.2023.154400. Available from: https://consensus.app/papers/details/4733b7c429f551c78b146d66cabf31f8/?utm_source
22. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med. 2021;181(8):1065-70. doi:10.1001/jamainternmed.2021.2626. Available from: https://consensus.app/papers/details/76c30a01075c576a8e8dddc245ff66be/?utm_source
23. Henry KE, Adams R, Parent C, Soleimani H, Sridharan A, Johnson L, et al. Factors driving provider adoption of the TREWS machine learning-based early warning system and its effects on sepsis treatment timing. Nat Med. 2022;28(7):1447-54. doi:10.1038/s41591-022-01895-z. Available from: https://consensus.app/papers/details/6d612895ceab5a858d0bdf45444741bc/?utm_source
24. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med. 2020;46(3):383-400. doi:10.1007/s00134-019-05872-y. Available from: https://consensus.app/papers/details/222e37db0df65cd8b05bb52638ea2620/?utm_source
25. Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early Prediction of Sepsis in the ICU Using Machine Learning: A Systematic Review. Front Med (Lausanne). 2021;8:607952. doi:10.3389/fmed.2021.607952. Available from: https://consensus.app/papers/details/33b12c07aeec5f1ba13767b6400b861f/?utm_source
26. Wang Z, Wang W, Sun C, Li J, Xie S, Xu J, et al. A methodological systematic review of validation and performance of sepsis real-time prediction models. NPJ Digit Med.

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

2025;8(1):190. doi:10.1038/s41746-025-01587-1. Available from: https://consensus.app/papers/details/2a57db1287765afa8f85b0bce101191d/?utm_source

27. Zhang Z, Chen L, Xu P, Wang Q, Zhang J, Chen K, et al. Effectiveness of automated alerting system compared to usual care for the management of sepsis. NPJ Digit Med. 2022;5(1):101. doi:10.1038/s41746-022-00650-5. Available from: https://consensus.app/papers/details/85d125e1cf15539fb950c3fbfbc97c49/?utm_source
28. Kim HJ, Ko RE, Lim SY, Park S, Suh GY, Lee YJ. Sepsis Alert Systems, Mortality, and Adherence in Emergency Departments. JAMA Netw Open. 2024;7(7):e2422823. doi:10.1001/jamanetworkopen.2024.22823. Available from: https://consensus.app/papers/details/f806798279595fcaa454eb009e1207ab/?utm_source
29. Warttig S, Alderson P, Evans DJ, Lewis SR, Kourbeti IS, Smith AF. Automated monitoring compared to standard care for the early detection of sepsis in critically ill patients. Cochrane Database Syst Rev. 2018;6:CD012404. doi:10.1002/14651858.cd012404.pub2. Available from: https://consensus.app/papers/details/2dfddbb8da6c5fbaa630bd8332196206/?utm_source
30. Adams R, Henry KE, Sridharan A, Soleimani H, Zhan A, Rawat N, et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. Nat Med. 2022;28(7):1455-60. doi:10.1038/s41591-022-01894-0. Available from: https://consensus.app/papers/details/abee295d872555919416a3ed8d23e5ba/?utm_source
31. Kamran F, Tjandra D, Heiler A, Virzi J, Singh K, King JE, et al. Evaluation of Sepsis Prediction Models before Onset of Treatment. NEJM AI. 2024;1(3):AIoa2300032. doi:10.1056/aioa2300032. Available from: https://consensus.app/papers/details/ad5c4d4a32305b8091e1deb47486e2bc/?utm_source
32. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLoS Med. 2021;18(3):e1003583. doi:10.1371/journal.pmed.1003583. Available from: https://consensus.app/papers/details/2525b0f4be60541ea19b8a79e9698d2c/?utm_source

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

33. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898. doi:10.1136/bmj.l4898. Available from: https://consensus.app/papers/details/99eca47c80a253ec80ec043138089451/?utm_source
34. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919. doi:10.1136/bmj.i4919. Available from: https://consensus.app/papers/details/87e695ba99c35aea9a306d7410d8d111/?utm_source
35. 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(1):51-8. doi:10.7326/m18-1376. Available from: https://consensus.app/papers/details/17a2602fae335f9b81b077edee90cd0c/?utm_source
36. Röver C, Knapp G, Friede T. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. BMC Med Res Methodol. 2015;15(1):99. doi:10.1186/s12874-015-0091-1. Available from: https://consensus.app/papers/details/dc41b84af1b45432aa6ae01d5100bcdf/?utm_source
37. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378. Available from: https://consensus.app/papers/details/34463dae28385dcf915dfb79a563a292/?utm_source
38. Snell KIE, Levis B, Damen JAA, Dhiman P, Debray TPA, Hooft L, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). BMJ. 2023;381:e073538. doi:10.1136/bmj-2022-073538. Available from: https://consensus.app/papers/details/8b8825b23cfc5028b90baa2fd73aab02/?utm_source
39. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol.

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

2014;14(1):135. doi:10.1186/1471-2288-14-135. Available from: https://consensus.app/papers/details/8123cbd49f0a5a5795aa835a99387201/?utm_source

40. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res. 2018;27(6):1785-805. doi:10.1177/0962280216669183. Available from: https://consensus.app/papers/details/543dfeddf55853b485db2e6f17936562/?utm_source
41. 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. 2019;28(9):2768-86. doi:10.1177/0962280218785504. Available from: https://consensus.app/papers/details/18bc42d9f1aa592795c759a0ed58e9f8/?utm_source
42. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505. doi:10.1136/bmj-2024-082505. Available from: https://consensus.app/papers/details/a5c81c0c866c5d92bd13b33ac4b45e5a/?utm_source
43. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol. 2014;14(1):25. doi:10.1186/1471-2288-14-25. Available from: https://consensus.app/papers/details/cb0f8813c7555c329bde150d9ac3e847/?utm_source
44. Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;356:i6460. doi:10.1136/bmj.i6460. Available from: https://consensus.app/papers/details/d2569a44fc8e52d8a38609f9f51797f4/?utm_source
45. Snell KI, Ensor J, Debray TP, Moons KG, Riley RD. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C -statistic and calibration measures?. Stat Methods Med Res. 2018;27(11):3505-

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

22. doi:10.1177/0962280217705678. Available from: https://consensus.app/papers/details/6c9a24bdc7b45677aed26572fcf42daa/?utm_source

46. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. J Clin Epidemiol. 2011;64(12):1294-302. doi:10.1016/j.jclinepi.2011.03.017. Available from: https://consensus.app/papers/details/5ba9f5bf7a3758618b8f2f9e1c475c04/?utm_source
47. Sterne JAC, Sutton AJ, Ioannidis JPA, 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. doi:10.1136/bmj.d4002. Available from: https://consensus.app/papers/details/37fc12e758405694a8068a87d7f3a59d/?utm_source
48. Shi L, Lin L. The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. Medicine (Baltimore). 2019;98(23):e15987. doi:10.1097/md.0000000000015987. Available from: https://consensus.app/papers/details/bf78dbb9ad0e5bd6a5f4a58c89e2b278/?utm_source
49. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64(4):401-6. doi:10.1016/j.jclinepi.2010.07.015. Available from: https://consensus.app/papers/details/3af999c5954f59dbaa499a9cffe96512/?utm_source
50. Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. J Clin Epidemiol. 2020;122:142-52. doi:10.1016/j.jclinepi.2019.12.021. Available from: https://consensus.app/papers/details/d870086163fd584bba27a387bce9ae91/?utm_source
51. Downing NL, Rolnick J, Poole SF, Hall E, Wessels AJ, Heidenreich P, et al. Electronic health record-based clinical decision support alert for severe sepsis: a randomised evaluation. BMJ Qual Saf. 2019;28(9):762-8. doi:10.1136/bmjqs-2018-008765. Available from: https://consensus.app/papers/details/477f2501763850029104d168dd1d42ca/?utm_source

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

52. Tarabichi Y, Cheng A, Bar-Shain D, McCrate BM, Reese LH, Emerman C, et al. Improving Timeliness of Antibiotic Administration Using a Provider and Pharmacist Facing Sepsis Early Warning System in the Emergency Department Setting: A Randomized Controlled Quality Improvement Initiative*. Crit Care Med. 2022;50(3):418-27. doi:10.1097/ccm.0000000000005267. Available from: https://consensus.app/papers/details/bf41330b0f6d597283e1dff927d6d37a/?utm_source
53. Boussina A, Shashikumar SP, Malhotra A, Owens RL, El-Kareh R, Longhurst CA, et al. Impact of a deep learning sepsis prediction model on quality of care and survival. NPJ Digit Med. 2024;7(1):14. doi:10.1038/s41746-023-00986-6. Available from: https://consensus.app/papers/details/ecc5a3a4bd2e5e488bf7887aae9e4899/?utm_source
54. Wong A, Currey D, Schwinne M, Park-Egan B, Meyer S, Gutting A, et al. Multicenter Prospective Validation of an Updated Proprietary Sepsis Prediction Model. JAMA Netw Open. 2026;9(2):e260181. doi:10.1001/jamanetworkopen.2026.0181. Available from: https://consensus.app/papers/details/5c1a6db4056952b1b9a581537e714a80/?utm_source
55. Hooper MH, Weavind L, Wheeler AP, Martin JB, Gowda SS, Semler MW, et al. Randomized trial of automated, electronic monitoring to facilitate early detection of sepsis in the intensive care unit*. Crit Care Med. 2012;40(7):2096-101. doi:10.1097/ccm.0b013e318250a887. Available from: https://consensus.app/papers/details/90c30fd65cec5eaaa754fe4d34cd9e8b/?utm_source
56. Semler MW, Weavind L, Hooper MH, Rice TW, Gowda SS, Nadas A, et al. An Electronic Tool for the Evaluation and Treatment of Sepsis in the ICU. Crit Care Med. 2015;43(8):1595-602. doi:10.1097/ccm.0000000000001020. Available from: https://consensus.app/papers/details/47c38e19eabc56b589b9079267af926a/?utm_source
57. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res. 2017;4(1):e000234. doi:10.1136/bmjresp-2017-000234. Available from: https://consensus.app/papers/details/a1d8abd75d6f577484338708a69aa5d4/?utm_source

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

58. McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Qual. 2017;6(2):e000158. doi:10.1136/bmjoq-2017-000158. Available from: https://consensus.app/papers/details/98e49aa123595a2287f111407c434d67/?utm_source
59. Burdick H, Pino E, Gabel-Comeau D, McCoy A, Gu C, Roberts J, et al. Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: a prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals. BMJ Health Care Inform. 2020;27(1):e100109. doi:10.1136/bmjhci-2019-100109. Available from: https://consensus.app/papers/details/3f2f7a79a5e45250ad8756bd0cb2e623/?utm_source
60. Cull J, Brevetta R, Gerac J, Kothari S, Blackhurst D. Epic Sepsis Model Inpatient Predictive Analytic Tool: A Validation Study. Crit Care Explor. 2023;5(7):e0941. doi:10.1097/cce.0000000000000941. Available from: https://consensus.app/papers/details/d82c6448284252debfa617758441325d/?utm_source
61. Zhang R, Long F, Zhao Z, Wu J, Tan R, Xu W, et al. Machine learning predicts sepsis deterioration trajectories. NPJ Digit Med. 2026;9(1):385. doi:10.1038/s41746-026-02565-x. Available from: https://consensus.app/papers/details/15222238068a5af1aa0ab3f060b71f01/?utm_source
62. Jung AD, Baker J, Droege CA, Nomellini V, Johannigman J, Holcomb JB, et al. Sooner is better: use of a real-time automated bedside dashboard improves sepsis care. J Surg Res. 2018;231:373-9. doi:10.1016/j.jss.2018.05.078. Available from: https://consensus.app/papers/details/14ce6074b0d35e46a6c5554a0e8f9efb/?utm_source
63. Chanas T, Volles D, Sawyer R, Mallow-Corbett S. Analysis of a new best-practice advisory on time to initiation of antibiotics in surgical intensive care unit patients with septic shock. J Intensive Care Soc. 2019;20(1):34-9. doi:10.1177/1751143718767059. Available from: https://consensus.app/papers/details/acf6cd191e0b5ed4ae8032ed54d1b83d/?utm_source
64. Dewan M, Vidrine R, Zackoff M, Paff Z, Seger B, Pfeiffer S, et al. Design, Implementation, and Validation of a Pediatric ICU Sepsis Prediction Tool as Clinical Decision Support. Appl

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

Clin Inform. 2020;11(02):218-25. doi:10.1055/s-0040-1705107. Available from: https://consensus.app/papers/details/c59689e1e8e35e7a89c7a5f6f83d3fee/?utm_source

65. Lipatov K, Daniels CE, Park JG, Elmer J, Hanson AC, Madsen BE, et al. Implementation and evaluation of sepsis surveillance and decision support in medical ICU and emergency department. Am J Emerg Med. 2022;51:378-83. doi:10.1016/j.ajem.2021.09.086. Available from: https://consensus.app/papers/details/3e50e2652c09531a8704905307ddbd7a/?utm_source
66. Alturki A, Al-Eyadhy A, Alfayez A, Bendahmash A, Aljofan F, Alanzi F, et al. Impact of an electronic alert system for pediatric sepsis screening a tertiary hospital experience. Sci Rep. 2022;12(1):12436. doi:10.1038/s41598-022-16632-2. Available from: https://consensus.app/papers/details/1fd7241a9cb95adba35f1350dff00e8e/?utm_source
67. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open. 2018;8(1):e017833. doi:10.1136/bmjopen-2017-017833. Available from: https://consensus.app/papers/details/683b845699a95a61be72b08f71d04eff/?utm_source
68. Barton C, Chettipally U, Zhou Y, Jiang Z, Lynn-Palevsky A, Le S, et al. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. Comput Biol Med. 2019;109:79-84. doi:10.1016/j.compbiomed.2019.04.027. Available from: https://consensus.app/papers/details/0adf35ddae6659f4bef86e4751ebfc2f/?utm_source
69. Burdick H, Pino E, Gabel-Comeau D, Gu C, Roberts J, Le S, et al. Validation of a machine learning algorithm for early severe sepsis prediction: a retrospective study predicting severe sepsis up to 48 h in advance using a diverse dataset from 461 US hospitals. BMC Med Inform Decis Mak. 2020;20(1):276. doi:10.1186/s12911-020-01284-x. Available from: https://consensus.app/papers/details/c7b5b99b923a50c8bd5c5764a1fead1f/?utm_source
70. Wang D, Li J, Sun Y, Ding X, Zhang X, Liu S, et al. A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients. Front Public Health. 2021;9:754348. doi:10.3389/fpubh.2021.754348. Available from: https://consensus.app/papers/details/53179252f0665f68b19dc9dc4a6ad06d/?utm_source

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

71. Moor M, Bennett N, Plečko D, Horn M, Rieck B, Meinshausen N, et al. Predicting sepsis using deep learning across international sites: a retrospective development and validation study. EClinicalMedicine. 2023;62:102124. doi:10.1016/j.eclinm.2023.102124. Available from: https://consensus.app/papers/details/bb649b32c6295503a439ae7e9bef417e/?utm_source
72. Kim T, Tae Y, Yeo HJ, Jang JH, Cho K, Yoo D, et al. Development and Validation of Deep-Learning-Based Sepsis and Septic Shock Early Prediction System (DeepSEPS) Using Real-World ICU Data. J Clin Med. 2023;12(22):7156. doi:10.3390/jcm12227156. Available from: https://consensus.app/papers/details/452ab54a0a8954aca04555c9ddeaa5e7/?utm_source
73. Dutta S, McMurry R, Tasi MC, Dunham L, McEvoy DS, Stump T, et al. Performance of a Sepsis Prediction Model Across Different Sepsis Definitions. JAMA Netw Open. 2026;9(4):e265599. doi:10.1001/jamanetworkopen.2026.5599. Available from: https://consensus.app/papers/details/746252201e985da9bb6b14a7f8b4787e/?utm_source
74. Joshi M, Ashrafian H, Arora S, Khan S, Cooke G, Darzi A. Digital Alerting and Outcomes in Patients With Sepsis: Systematic Review and Meta-Analysis. J Med Internet Res. 2019;21(12):e15166. doi:10.2196/15166. Available from: https://consensus.app/papers/details/a8074edec18b5b76b02277b78ef5c0d9/?utm_source
75. Patel H, Crusco S, Hansen D, Morales J, Patale T, Chiuzan C, et al. A Systematic Review and Meta-analysis of Externally Validated Epic Clinical Decision Support Tools. J Gen Intern Med. 2026. Published online ahead of print. doi:10.1007/s11606-026-10381-y. Available from: https://consensus.app/papers/details/a87417f01bff53ed9c9011fccd861bd8/?utm_source
76. Abbas GH, Sen P, Giri OA, Khan NH. Artificial Intelligence-Based Predictive Modeling for Early Detection of Sepsis in Hospitalized Patients: A Systematic Review and Meta-Analysis. Crit Care Explor. 2025;7(12):e1360. doi:10.1097/cce.0000000000001360. Available from: https://consensus.app/papers/details/e43d74e66e1551449aa3afad08cff47c/?utm_source
77. Austrian JS, Jamin CT, Doty GR, Blecker S. Impact of an emergency department electronic sepsis surveillance system on patient mortality and length of stay. J Am Med Inform Assoc. 2018;25(5):523-9. doi:10.1093/jamia/ocx072. Available from: https://consensus.app/papers/details/acee3cf40d8b50abb26661dc4af83f1b/?utm_source

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved

78. Ranard BL, Qian M, Cummings MJ, Zhang DY, Lee SM, Beitler JR, et al. Pragmatic, multicentre, factorial, randomised controlled trial of sepsis electronic prompting for timely intervention and care (SEPTIC trial): a protocol. BMJ Open. 2025;15(8):e088792. doi:10.1136/bmjopen-2024-088792. Available from: https://consensus.app/papers/details/e67a2f6fe9805c81991aa35bed32d2a5/?utm_source
79. Morrison E, Overley J, Wurm K, Mock D, Bhavani SV. Development to implementation to evaluation: step-by-step approach to embedding a clinical trial of a machine learning algorithm in the electronic health record. JAMIA Open. 2026;9(5):ooag180. doi:10.1093/jamiaopen/ooag180. Available from: https://consensus.app/papers/details/af8c35616a5c5fb4bd64c3393a8296d3/?utm_source

Downloaded from theInternationalmedicaljournal.org. For personal use only. No other uses without permission.
Copyright © International Medical Journal Corp. Ltd. All rights reserved