Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Design and applicability of clinical prediction models for serious infection in children with fever: a scoping review using causal directed acyclic graphs

Ariel O Mace
ariel.mace@health.wa.gov.au

Perth Children's Hospital
Andrew C Martin
Perth Children's Hospital
Peter Richmond
The Kids Research Institute Australia
Thomas L Snelling
The Kids Research Institute Australia
Yue Wu
The University of Sydney

Research Article

Keywords: clinical prediction rules, bacterial infections, paediatrics, causal directed acyclic graphs, knowledge translation

Posted Date: August 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10274735/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Clinical prediction models (CPMs) have been developed to estimate the risk of serious or invasive bacterial infection (SBI/IBI) in febrile children and support clinical decision-making, however routine clinical uptake remains limited. Differences in study design and model development may hinder comparison of available tools and assessment of suitability for specific contexts. This scoping review aimed to characterise the key design and development features of available CPMs and map them within causal directed acyclic graphs (DAGs) to examine how they may influence transportability and clinical applicability across clinical settings.

Methods

A scoping review was conducted per the JBI framework. Medline, Embase and relevant bibliographies were systematically searched to identify CPMs predicting SBI/IBI risk in febrile children (< 18y) presenting to emergency and/or ambulatory care. Extracted data included study design, derivation cohort characteristics, outcome definitions, predictors, modelling approaches and output types. DAGs were constructed to depict hypothesised causal dependencies between key concepts.

Results

Of 4010 studies identified, 56 met inclusion criteria. Substantial heterogeneity was observed across all aspects of study design and model development. Most CPMs were developed in populations aged under 3 months (n = 35), however age-ranges and cohort selection criteria were diverse. SBI was the predominant target outcome (n = 36), however definitions differed and reported prevalence varied substantially (0.8%-37.4%). Developed models varied in algorithm complexity, predictor selection and output format, with potential implications for interpretability and usability. Algorithms included straightforward rule-based, decision-tree, and score-based models, and more complex statistical and machine-learning approaches. Predictor requirements differed in number (2–28 predictors/model), type, definition and handling of missing data. Prediction outputs were predominantly categorical risk-groups intended to guide clinical management. Models generating probabilistic risk estimates often demonstrated better discrimination but did not consistently define clinically-actionable decision thresholds. Constructed DAGs illustrated pathways through which cohort selection and outcome definitions may influence case ascertainment, predictor-outcome relationships, and model transportability.

Conclusions

CPMs for predicting SBI/IBI in febrile children demonstrate substantial heterogeneity extending beyond predictive performance alone. Causal DAGs can provide a transparent framework for comparing and evaluating the transportability and applicability of available CPMs, highlighting opportunities to align core design elements to improve comparability, implementation and clinical utility.

Clinical Trial Registration (if any)

: Not applicable

Introduction

Fever in children presents a common diagnostic challenge for clinicians.[1] Most febrile illnesses are attributable to self-limiting infections that are typically caused by viruses and resolve with supportive measures only. However, a small proportion are caused by invasive bacterial infections (IBIs), such as bacteraemia and meningitis, or other serious bacterial infections (SBIs) including urinary tract infection (UTI), that carry a significant risk of morbidity and mortality if diagnosis and treatment are delayed.[2–4] While the management of those with an identifiable infective focus or who are very unwell is usually straightforward, many children, especially young infants, present with non-specific symptoms making the underlying cause of fever unclear.[3,5] No individual clinical feature or laboratory biomarker has yet been shown to reliably rule in or rule out serious infection in febrile children with sufficient certainty to guide clinical decision-making;[6,7] clinicians must therefore integrate multiple, often non-specific, clinical findings to infer the likelihood of serious infection, a process that can be cognitively challenging without support. Resultant diagnostic uncertainty can lead to both under-recognition of serious infections, with delayed management and poor patient outcomes, and over-investigation and treatment of those with self-limiting illnesses, with inherent risks of iatrogenic harm, promotion of antimicrobial resistance and unwarranted burden on families and health services.[8]

Clinical prediction models (CPMs), which combine two or more patient-specific variables (such as demographic, history, examination or laboratory characteristics) to estimate the probability of the presence (diagnosis) of a specific outcome, have therefore been developed to support this decision-making process.[9,10] Many developed CPMs have demonstrated good predictive performance for identifying febrile children presenting to emergency departments (ED) or ambulatory care who are at either 'high-risk' or 'low-risk' of IBI or SBI,[6,11–13] and have been shown to improve diagnostic testing, antibiotic prescribing and health-care utilisation for children with fever.[14–16] However, despite increasing interest in CPM development and evidence of their potential benefits, their uptake into routine clinical practice remains limited.[17–19] While predictive performance is often the primary focus of model evaluation, successful CPM implementation also depends on its applicability and feasibility for use within the target setting, based on factors including population similarity and

epidemiological context, predictor availability, outcome definitions, and compatibility for use within local healthcare systems.[20–22] Previous reviews have highlighted considerable heterogeneity in the design of these available CPMs in ways which may limit a clinician's ability to compare models, assess their generalisability, and determine suitability for their context.[6,13,23] Furthermore, increasing use of more complex statistical and machine-learning approaches may have consequences for model interpretability, resulting in 'black-box' predictions that affect explainability, usability and, thereby, clinician trust and willingness to implement the tool.[20,22,24] Even a CPM with excellent predictive performance is unlikely to be adopted if clinicians do not perceive it to be applicable to their clinical setting, or if prediction methods and outputs are not presented in a way that is deemed transparent, interpretable and usable.[9,25,26]

Continued development of CPMs with limited clinical uptake represents a missed opportunity to impact on patient care and an inefficient use of research resources.[27–29] Equally, the use of CPMs outside of the clinical contexts for which they were developed may lead to inaccurate predictions and potential patient harm.[30] Selecting an appropriate model requires clinicians to understand not only the differences between available models, but also how these differences may influence their applicability, transportability and expected performance in their own clinical setting.[20] To date, published reviews of CPMs for febrile children have focused primarily on evaluating diagnostic performance, providing only broad characterisation of the methodological features underlying model heterogeneity, and have often been restricted to young infant populations.[6,11–13,31,32] One review[23] more directly compared the study and model design characteristics of CPMs identified in two earlier systematic reviews of individual predictors,[31,32] identifying features that may limit diagnostic research in febrile children and recommending greater standardisation of definitions and outcomes to improve comparability and uptake. While providing valuable insights into sources of heterogeneity, further detailed characterisation of individual model features is needed to support comparison of available CPMs and assessment of their applicability and transportability across clinical settings. A structured, transparent framework for this process may complement existing reviews and support more informed model selection, implementation, and future development.[20,33]

Qualitative causal diagrams, in the form of directed acyclic graphs (DAGs), are ideally suited to facilitate this evaluation and are now recommended to guide the design and analysis of clinical studies.[34,35] Causal DAGs are graphical representations of hypothesised causal networks, in which variables relevant to a domain are represented as nodes and connected by directed arrows indicating their direct causal effects,[36] providing a transparent framework for understanding and evaluating how methodological choices may influence study and model outcomes.[35,37,38] This scoping review therefore aimed to systematically identify and compare key study and model design characteristics of published CPMs developed to predict SBI and/or IBI risk in febrile children, and to map these concepts within causal DAGs to examine how model applicability and performance is likely to be influenced by:

1. model training cohort selection and its alignment with the target population;
2. prediction target definition and its alignment with outcomes of clinical importance;
3. model algorithm type and how it is explained and presented to the end-user;
4. model predictor variable selection and its relationship to usability; and
5. what and how model outputs are presented to the end-user.

Methods

Study identification

A scoping review methodology was selected as it aligns with questions focused on mapping the extent, range, and characteristics of an evidence base, consistent with the objectives of this review, and was performed in accordance with the JBI framework.[39] Medline and Embase databases were systematically searched on 7 April 2024 and updated 14 May 2025 to identify studies reporting on the development (derivation) and/or evaluation of one or more original diagnostic CPM (also referred to as a clinical decision rule, tool, model, support system, index, score or algorithm)[20] estimating the risk or probability of SBI and/or IBI in febrile children less than 18 years old presenting to ED and/or ambulatory care.

The search was limited to English language articles with full text available; full search terms are listed in Supplementary material A. Studies were excluded if reporting on (i) CPMs estimating prognosis or intervention outcomes only, (ii) CPMs predicting infection of a specific body location (e.g. meningitis), pathogen (e.g. malaria), clinical syndrome (e.g. sepsis) or hospital-acquired infection (e.g. line-related) only, (iii) CPMs intended only for newborns prior to hospital separation or children with pre-existing high-risk conditions (e.g. oncological conditions), (iv) CPMs developed for use in low-resource settings, (v) CPMs developed using novel/experimental markers, (vi) re-validation, implementation and impact studies, unless describing new model development or updates to pre-existing models, and (vii) review articles, letters/editorials, local guidelines, case reports, study protocols, dissertations, books, and articles not accessible after using library resources.

Titles, abstracts and full texts of identified studies were independently reviewed by two investigators (AOM, YW). Articles were excluded based on abstract if clearly meeting any exclusion criteria. Full texts of relevant abstracts were screened for eligibility and additional articles were identified through manual bibliography review of relevant studies.

Data extraction and analysis

Data were systematically extracted into a standardised Excel data collection form by investigator AOM. Selected data items were based on the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis checklist.[20] Extracted study characteristics included author(s), publication year, country(ies), clinical setting(s), study training cohort selection criteria and size, and prediction target outcome characteristics, including diagnostic criteria. *Prediction target outcome* was defined as the composite group of site-specific bacterial infections for risk prediction by the CPM, including

(but not limited to) any combination of bacteraemia, UTI, skin and soft tissue infection (SSTI), respiratory tract infection, intracranial infection, and bone and joint infection (BJI), or bacteraemia in isolation. *Diagnostic criteria* was defined as the study criteria used as the basis for judging the 'true' presence or absence of a site-specific infection.

Extracted model-specific data included algorithm presentation type and derivation methods, predictor variable characteristics, prediction output type, predictive performance, and ability to generate a prediction if input data were missing. If the development of more than one model was reported, characteristics of the study's principal model were extracted. Predictor *variables* were defined as patient-specific features assigned a value or weight within the model to inform predictions and were categorised *a priori* into 'background factors' (patient demographics, past medical history), 'history of presenting complaint' (reported symptoms), 'examination signs' (observable features) or 'investigation results'. *Prediction output type* was defined as how risk estimates were presented to the end-user. Measures evaluating the model's accuracy in distinguishing between those with or without the target outcome (*predictive performance*) included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and/or area under the receiver operating characteristic curve (AUROC). If not reported, performance characteristics were calculated using 2x2 tables constructed from available data when possible; study authors were not contacted for unpublished data.

Data were analysed thematically and presented as a narrative, with results synthesised in tables, figures or mapped within qualitative causal DAGs. A meta-analysis of performance characteristics was not planned due to the anticipated heterogeneity of included CPMs. DAGs were constructed by authors AOM and YW using GeNIe software (https://bayesfusion.com/downloads/). All DAG variables (**bolded** within the descriptive text) were colour-coded, labelled with a letter ('s' for summary nodes or 'd' for individual characteristic nodes) and consecutively numbered (e.g. d1, d2…), and defined within a variable dictionary (Supplementary material B). An illustrative high-level DAG is shown in Fig. 1, depicting hypothesised causal influences (either positive or negative) between **background factors (s1)** (blue nodes), **symptoms and signs (s3)** (pink nodes), **investigation results (s6)** (orange nodes), **management decisions/interventions (s5)** (green nodes), **operational factors (s5)** (yellow nodes), **diagnostic outcomes (s7)** (purple nodes), and the latent (or unobservable) **infection status and host response (s2)** (white nodes). As DAGs cannot represent feedback loops, arrow direction was assigned according to the predominant direction of influence relevant to the network's intended purpose. If required, a time series was depicted by duplicated nodes using 't +/- 1' (i.e. prior to or after the depicted variable).

Ethics and consumer involvement

Human research ethics committee approval was not required. While consumers were not directly involved in the conduct of this review, it was performed within a broader research program informed and endorsed by consumer representatives from an Infectious Disease Community Reference Group.

Results

Study setting

Of 4010 studies identified, 56 original studies published between 1985–2025 were included (Fig. 2; Fig. 3). Most studies were conducted in the USA (n = 21), the Netherlands (n = 7) or Spain (n = 4) (Table 1). Fifty-two studies were performed within paediatric or mixed adult/paediatric EDs, three in both emergency and ambulatory care,[40–42] and one in ambulatory care only[43]. Cohort size ranged from 52 to 16,967 participants. Figure 4 illustrates how **geographical (d4), temporal (d5)** and **clinical setting (d6)** characteristics may relate to the underlying **presence of target infection/s (d9)** among children included in a study cohort.

Table 1
Study and model characteristics by target prediction outcome category and age group criteria

Study publication details	Model training cohort characteristics				Model design characteristics		Predictor variables by type, n		
Author, publication year	**Setting, country**	**Age range/ fever definition**	**Other exclusion criteria**	**No. with outcome/ total cohort, n/N (%)**	**Algorithm type/output**	**Total variable no.**	**BG**	**HPC**	**Exam**
Bacteraemia only									
Broner (1990)[41]	ED/Amb care, USA	≤8 wks/ ≥38.1°C	Nil	5/52 (9.6%)	Rule/risk category	3			Appearance (compo
Crain (1982)[91]	ED, USA	≤8 wks/ ≥38.0°C	Focal signs	5/1982 (0.3%)	Rule/risk category	3			Appearance (compo
Gómez (2010) [92]	ED, Spain	< 90 d/ ≥38.0°C	Focal signs	23/101 (2.3%)	SDT/risk category	6	H/o comorb		Appearance (compo
Tsai (2020)[93]	ED, Taiwan	< 18 yrs/ NS	Nil	146/16,967 (0.9%)	ML/risk category	10	Age		
Isaacman (2000)[56]	ED, USA	3–36 mo/ NS	H/o comorb, recent a/bs	46/633 (7.3%)	SM/prob estimate	3	Gender		Max temp
Crocker (1985)[52]	ED, USA	6 mo–2 yrs/ ≥39.4°C	'unwell'	21/201 (10.4%)	Rule/risk category	2		Fever duration	
IBI (defined as bacteraemia and bacterial meningitis)									
Yin (2022)[53]	ED, China	≤28 d/ ≥38.0°C	H/o prem, h/o comorb, 'unwell', recent a/bs, 'other'	166/1053 (15.8%)	SDT/risk category	7	Age		Abn neurology, appe (composite)
Berkowitz (1985)[40]	ED/Amb care, USA	< 2 mo/ ≥38.0°C	Focal signs	15/434 (3.5%)	Rule/risk category	6	Age, breastfeeding	Sick contacts	Lethargy/alertness
Yang (2023)[62]	ED, Taiwan	≤60 d/ ≥38.0°C	H/o comorb	53/184 (2.9%)	ML/not specified	5	Age		Max temp, heart rate/tachycardia
Aronson (2019)[48]	ED, USA	≤60 d/ ≥38.0°C	H/o comorb, 'unwell'	NA (case control)	Score/risk category	4	Age		Max temp
Velasco (2015)[94]	ED, Spain	< 90 d/ ≥38.0°C	Focal signs	50/766 (6.5%)	SDT /risk category	5	Age		Appearance (compo
Yaeger (2022)[63]	ED, USA	≤90 d/ ≥38.0°C	Nil	39/2311 (1.7%)	ML/not specified	9	Age, h/o comorb, gender, gestational age	Fever duration, illness duration	Appearance, cough
Mintegi (2014)[51]	ED, Spain/Italy	< 3 mo/ ≥38.0°C	Focal signs	48/1123 (4.3%)	SDT/risk category	6	Age		Appearance (compo
Pantell (2004)[43]	Amb care, USA	≤3 mo/ ≥38.0°C	H/o prem, h/o comorb	63/3066 (2.1%)	SDT/risk category	3	Age		Appearance, max te
Ballard (2024)[44]	ED, USA	7–60 d/ ≥38.0°C	Nil	196/4411 (4.4%)	ML/prob estimate	25	Age, gestational age, h/o sibling with GBS, perinatal a/bs (2 items), h/o comorb, h/o		Max temp

Abbreviations: a/bs antibiotics; Amb Ambulatory care; ANC absolute neutrophil count; BG background factor; comorb medical comorbities/complexitiesCRP cerebrospinal fluid; CXR chest X-ray; d days old; ED emergency department (paediatric and/or mixed adult/paediatric); ENT ear nose throat; ESR erythrocyte examination finding; GBS group B *Streptococcus*; Hb haemoglobin; h/o history of; HPC history of presenting complaint feature; IBI invasive bacterial infectio function tests; max maximum; ML machine-learned; mo months old; NA not applicable; NS not specified; neut neutrophil; PCT procalcitonin; PMN polymorpl SaO2 oxygen saturations; SBI serious bacterial infection; SDT sequential decision tree; SM statistical model; WCC white cell count wks weeks old; yrs years

Study publication details	Model training cohort characteristics				Model design characteristics		Predictor variables by type, n		
Author, publication year	Setting, country	Age range/ fever definition	Other exclusion criteria	No. with outcome/ total cohort, n/N (%)	Algorithm type/output	Total variable no.	BG	HPC	Exam
Bacteraemia only									
							neonatal fever, other non-specified (7 items)		
Nguyen (2021)[95]	ED, USA	7–60 d / ≥38.0°C	Focal signs, h/o prem, h/o comorb, 'unwell', 'other'	67/1803 (3.7%)	SDT/risk category	5	Age		Max temp
Burstein (2022)[96]	ED, Canada	8 - ≤60 d/ ≥38.0°C	Focal signs, h/o prem, h/o comorb	27/957 (2.8%)	Rule/risk category	3			Max temp
Burstein (2025)[97]	ED, Canada	8–60 d/ ≥38.0°C	Focal signs, h/o prem, h/o comorb, 'unwell', recent a/bs, 'other'	38/1987 (1.9%)	SDT/risk category	3			Max temp
Dorney (2022)[98]	ED, USA	31–60 d/ ≥38.0°C	Focal signs, h/o prem, h/o comorb, 'unwell', recent a/bs, 'other'	7/616 (1.1%)	Rule/risk category	7	Vaccination history	No recent a/bs	Appearance (compo signs
Feildel-Fournial (2023)[99]	ED, France	> 6 d - <5 yrs/ ≥38.0°C	Focal signs	11/1061 (1.0%)	SDT/risk category	6	Age, h/o comorb	Fever duration	Appearance (compo
SBI (defined as bacteraemia, bacterial meningitis and UTI only)									
Bonadio (1993)[68]	ED, USA	0–8 wks/ ≥38.0°C	'other'	29/233 (12.4%)	Score/risk category	3			Affect (e.g. smiling/irritability/c circulation/capillary respiratory (effort/ra difficulty)
Kuppermann (2019)[54]	ED, USA	≤60 d/ ≥38.0°C	H/o prem, h/o comorb, 'unwell', recent a/bs, 'other'	179/1821 (9.3%)	SDT/risk category	3			
Ramgopal (2020)[66]	ED, USA	≤60 d / ≥38.0°C	H/o prem, h/o comorb, 'unwell', recent a/bs, 'other'	138/1470 (9.4%)	ML/not specified	4			

Abbreviations: a/bs antibiotics; Amb Ambulatory care; ANC absolute neutrophil count; BG background factor; comorb medical comorbities/complexitiesCRP cerebrospinal fluid; CXR chest X-ray; d days old; ED emergency department (paediatric and/or mixed adult/paediatric); ENT ear nose throat; ESR erythrocyte examination finding; GBS group B *Streptococcus*; Hb haemoglobin; h/o history of; HPC history of presenting complaint feature; IBI invasive bacterial infection function tests; max maximum; ML machine-learned; mo months old; NA not applicable; NS not specified; neut neutrophil; PCT procalcitonin; PMN polymorph SaO2 oxygen saturations; SBI serious bacterial infection; SDT sequential decision tree; SM statistical model; WCC white cell count wks weeks old; yrs years

Study publication details	Model training cohort characteristics				Model design characteristics		Predictor variables by type, n		
Author, publication year	**Setting, country**	**Age range/ fever definition**	**Other exclusion criteria**	**No. with outcome/ total cohort, n/N (%)**	**Algorithm type/output**	**Total variable no.**	**BG**	**HPC**	**Exam**
Bacteraemia only									
Bachur (2001)[100]	ED, USA	≤90 d/ ≥38.0°C	Nil	373/5279 (7.1%)	SDT/risk category	4	Age		Max temp
Poirier (2021)[67]	ED, Canada	≤90 d/ NS	Focal signs, h/o prem, h/o comorb	442/3461 (12.7%)	Rule/risk category	4			
Yaeger (2021)[64]	ED, USA	≤90 d / ≥38.0°C	Nil	67/877 (7.6%)	ML/not specified	10	Age, h/o comorb, gender, gestational age, insurance	Illness duration	Appearance (compo cough, max temp
Chong (2023)[45]	ED, Singapore	< 3 mo/ ≥38.0°C	Nil	224/1002 (22.4%)	ML/prob estimate	6	Age, gender	Fever duration	Max temp
SBI (defined using varying infection combinations)									
Marom (2007)[101]	ED, Israel	< 28 d/ ≥38.0°C	Nil	108/386 (28.0%)	Rule/risk category	8	H/o comorb, h/o prematurity	No recent a/bs	Appearance, focal s
Li (2022)[102]	ED. Taiwan	< 30 d/ ≥38.0°C	H/o prem, h/o comorb, recent a/bs	38/131 (29.0%)	Score/prob estimate	4			Appearance (compo
Herr (2001)[49]	ED, USA	≤60 d / ≥38.0°C	Nil	41/404 (10.1%)	Rule/risk category	14	H/o comorb, perinatal a/bs, h/o prematurity, prior hospitalisations, sibling with GBS	No recent a/bs	Appearance, focal s
Vujevic (2017)[103]	ED, Turkey	< 90 d/ ≥38.0°C	H/o prem, h/o comorb, recent a/bs, 'other'	70/187 (37.4%)	SM/prob estimate	4	Gender		Max temp
Chang (2021)[46]	ED, Singapore	≤90 d/ ≥38.0°C	H/o prem, h/o comorb, recent a/bs	54/187 (28.9%)	SM/not specified	7	Age, gender	Fever duration	
Dagan (1985)[104]	ED, USA	< 3 mo/ NS	H/o prem, h/o comorb, recent a/bs	23/233 (9.9%)	Rule/risk category	4			Focal signs
Pinto (2017)[105]	ED, Spain	0–3 mo/ ≥38.0°C	Focal signs, h/o comorb	159/702 (22.6%)	SM/prob estimate	7	H/o comorb		Appearance
Olaciregui (2009)[106]	ED, Spain	4 – 90 d/ >38.0°C	Focal signs, h/o comorb, recent a/bs, 'other'	82/347 (23.6%)	Rule/risk category	5			Appearance

Abbreviations: a/bs antibiotics; Amb Ambulatory care; ANC absolute neutrophil count; BG background factor; comorb medical comorbities/complexitiesCRF cerebrospinal fluid; CXR chest X-ray; d days old; ED emergency department (paediatric and/or mixed adult/paediatric); ENT ear nose throat; ESR erythrocyte examination finding; GBS group B *Streptococcus*; Hb haemoglobin; h/o history of; HPC history of presenting complaint feature; IBI invasive bacterial infectio function tests; max maximum; ML machine-learned; mo months old; NA not applicable; NS not specified; neut neutrophil; PCT procalcitonin; PMN polymorpl SaO2 oxygen saturations; SBI serious bacterial infection; SDT sequential decision tree; SM statistical model; WCC white cell count wks weeks old; yrs years

Study publication details	Model training cohort characteristics				Model design characteristics		Predictor variables by type, n		
Author, publication year	**Setting, country**	**Age range/ fever definition**	**Other exclusion criteria**	**No. with outcome/ total cohort, n/N (%)**	**Algorithm type/output**	**Total variable no.**	**BG**	**HPC**	**Exam**
Bacteraemia only									
Baker (1993)[107]	ED, USA	29–56 d/ ≥38.2°C	H/o comorb	65/747 (8.7%)	Rule/risk category	7			Appearance (compo signs
Bonadio (1993)[108]	ED, USA	4–8 wks /≥38.0°C	H/o comorb	24/534 (4.5%)	Rule/risk category	6			Appearance (compo signs
Baskin (1992)[109]	ED, USA	28–89 d/ ≥38.0°C	Focal signs, h/o prem, h/o comorb, 'unwell', recent a/bs, 'other'	27/503 (5.4%)	Rule/risk category	7	Vaccination history	No recent a/bs	Appearance (compo signs
Leroy (2018)[61]	ED, Switzerland/ Canada	< 3 yrs/ NS	Focal signs, h/o comorb, recent a/bs, 'other'	211/877 (24.1%)	SDT/risk category	4	Age		
Lee (2022)[58]	ED, South Korea	< 6 yrs/ NS	Nil	466/11,973 (3.9%)	ML/not specified	18	Daycare, gender, vaccination history (3 items)	Fever duration	Max temp, heart rate/tachycardia, ra respiratory (effort, r difficulty)
Van den Bruel (2007)[42]	ED/Amb care, Belgium	0–16 yrs/ NS	Nil	31/3981 (0.8%)	SDT/risk category	5	Age	H/o diarrhoea	Gut feeling/sense o reassurance, max te respiratory (effort, r difficulty)
Vorwerk (2011)[65]	ED, UK	0–16 yrs/ NS	Nil	42/506 (8.3%)	Rule/risk category	3			
Irwin (2017)[110]	ED, UK	< 16 yrs/ >38.0°C	H/o comorb, 'other'	264/1101 (24.0%)	SM/prob estimate	5			Normal/abn chest s respiratory (effort, r difficulty)
Lacour (2008)[50]	ED, Switzerland	7 d–36 mo/ >38.0°C	Focal signs, h/o comorb, 'other'	54/202 (26.7%)	Score/risk category	3			
Berger (1996)[47]	ED, The Netherlands	2 wks–1 year/ ≥38.0°C	H/o prem, h/o comorb, recent a/bs, 'other'	33/138 (23.9%)	SDT/prob estimate	5		H/o diarrhoea, fever duration	Appearance (compo signs
Bleeker (2001)[111]	ED, The Netherlands	1–36 mo/ ≥38.0°C	Focal signs, h/o comorb	58/231 (25.1%)	Score/prob estimate	10	Age	Fever duration, urine output, h/o vomiting	Max temp, circulatio refill, respiratory (ef difficulty)
Bleeker (2007)[112]	ED, The Netherlands	1–36 mo/ ≥38.0°C	Focal signs, h/o comorb	99/381 (26.0%)	Score/prob estimate	8		Fever duration, h/o vomiting	Appearance, circulation/capillary respiratory (effort, r difficulty)
Roukema (2008)[70]	ED, The Netherlands	1–36 mo/ ≥38.0°C	Focal signs	138/595 (23.2%)	SM/prob estimate	7	Age	Fever duration,	Appearance, circulation/capillary

Abbreviations: a/bs antibiotics; Amb Ambulatory care; ANC absolute neutrophil count; BG background factor; comorb medical comorbities/complexitiesCRF cerebrospinal fluid; CXR chest X-ray; d days old; ED emergency department (paediatric and/or mixed adult/paediatric); ENT ear nose throat; ESR erythrocyte examination finding; GBS group B *Streptococcus*; Hb haemoglobin; h/o history of; HPC history of presenting complaint feature; IBI invasive bacterial infectio function tests; max maximum; ML machine-learned; mo months old; NA not applicable; NS not specified; neut neutrophil; PCT procalcitonin; PMN polymorpl SaO2 oxygen saturations; SBI serious bacterial infection; SDT sequential decision tree; SM statistical model; WCC white cell count wks weeks old; yrs years

Study publication details	Model training cohort characteristics				Model design characteristics		Predictor variables by type, n		
Author, publication year	**Setting, country**	**Age range/ fever definition**	**Other exclusion criteria**	**No. with outcome/ total cohort, n/N (%)**	**Algorithm type/output**	**Total variable no.**	**BG**	**HPC**	**Exam**
Bacteraemia only									
					and risk category			h/o vomiting	items), respiratory (difficulty)
Brent (2011)[69]	ED, UK	4 wks–15 yrs/ NS	Nil	74/1951 (3.8%)	Score/prob estimate	8	H/o comorb, developmental delay		Max temp, hydration membranes), hypox lethargy/alertness, circulation/capillary respiratory (effort, ra difficulty)
Nijman (2013)[60]	ED, The Netherlands	1 mo–15 yrs/ ≥38.0°C	H/o comorb, recent a/bs	341/2717 (12.6%)	SM/prob estimate	11	Age, gender	Fever duration	Appearance (compo temp, heart rate/tac hypoxia/SaO2, circulation/capillary respiratory (effort, ra difficulty (2 items)
Nijman (2014)[113]	ED, The Netherlands	1 mo–16 yrs/ ≥38.5°C	H/o comorb, 'other'	170/1084 (15.7%)	SM/prob estimate	4		Fever duration	
Nijman (2018)[114]	ED, The Netherlands	1 mo–16 yrs/ ≥38.5°C	H/o comorb, 'other'	171/1085 (15.8%)	SM/prob estimate	12	Age, gender	Fever duration	Appearance (compo temp, heart rate/tac hypoxia/SaO2, circulation/capillary respiratory (effort, ra difficulty (2 items)
Urbane (2022)[59]	ED, Latvia	1 mo - <18yrs/ ≥38.0°C	H/o comorb	138/517 (26.7%)	Score/risk category	6		Fluid intake	Normal/abn chest s feeling/sense of rea items), circulation/c refill, respiratory (eff difficulty)
Goh (2006)[115]	ED, Singapore	3 mo–36 mo/ ≥38.0°C	Focal signs, recent a/bs	17/86 (19.8%)	Rule/risk category	2		Fever duration	
Hagedoorn (2021)[57]	ED, 8 (Europe)	0–18 yrs/ ≥38.0°C	Focal signs	135/12,268 (1.1%)	SM/prob estimate	13	Age, h/o comorb, gender	Fever duration	Abn neurology, appe max temp, heart rate/tachycardia, hy rash, respiratory (eff difficulty)(2 items)
SBI (predicting 3 individual infections - UTI, bacteraemia, pneumonia)									
Craig (2010)[55]	ED, Australia	< 5 yrs/ ≥38.0°C	H/o comorb	1120/15,781 (7.1%)	SM/prob estimate	28	Age, h/o comorb, gender, vaccination history (2 items)	H/o diarrhoea, illness duration, fluid intake, h/o sick contact, urinary symptoms, h/o vomiting	Appearance, bulging chest crackles, coug ENT findings, focal s temp, heart rate/tac circulation/capillary normal/abn chest s respiratory (effort, ra difficulty)(2 items), r symptoms, stridor, v

Abbreviations: a/bs antibiotics; Amb Ambulatory care; ANC absolute neutrophil count; BG background factor; comorb medical comorbities/complexitiesCRP cerebrospinal fluid; CXR chest X-ray; d days old; ED emergency department (paediatric and/or mixed adult/paediatric); ENT ear nose throat; ESR erythrocyte examination finding; GBS group B *Streptococcus*; Hb haemoglobin; h/o history of; HPC history of presenting complaint feature; IBI invasive bacterial infection function tests; max maximum; ML machine-learned; mo months old; NA not applicable; NS not specified; neut neutrophil; PCT procalcitonin; PMN polymorph SaO2 oxygen saturations; SBI serious bacterial infection; SDT sequential decision tree; SM statistical model; WCC white cell count wks weeks old; yrs years

Model training cohorts

Selection criteria for model training cohort inclusion were heterogenous, utilising differing combinations of patient background factors (e.g. **age (d1)**, history of **medical comorbidity or complexity (d3)**), presenting signs and symptoms (e.g. **fever characteristics (d10)**, **evidence focal signs of infection (d11)** or **appearance on presentation (d12)**) and/or operational factors (e.g. **test result availability (d17)**) (Table 1; Fig. 4). Most CPMs were trained in populations aged under 3 months old (n = 35), with a total of 22 distinct age ranges (up to 18 years) identified. Temperature thresholds to define 'fever' were specified in 47 studies, ranging from ≥ 38.0° (n = 39) to ≥ 39.4°C (n = 1). When specified, accepted measurement methods included differing combinations of axillary (n = 8), rectal (n = 26) and/or oral (n = 2) measurement, and/or subjective report of fever (n = 2); seven studies included participants only if fever was recorded in

the clinical setting.[44–50] Twenty-four studies (43%) excluded otherwise eligible patients due to **test result (un)availability (d17)** or incomplete assessment data (Supplementary material C); four studies[51–54] reported excluding more than one quarter of otherwise eligible participants due to missing data. Figure 4 depicts potential pathways through which study selection criteria, common clinical management decisions/interventions performed either prior to or on presentation (e.g. recent **antipyretic or analgesic use (d8)**, **timing of antibiotic administration at presentation (d13)**, or **decision to perform investigations (d15)**) and underlying operational factors (e.g. **setting-specific factors (d14)** including local clinical management guidelines), may influence both participant inclusion and the underlying prevalence (pre-test probability) of infection (i.e. **presence of target infection/s (d9)**) within a study cohort.

Prediction target outcomes

Prediction target outcome categories included 'SBI' (n = 36), 'IBI' (n = 14) and 'bacteraemia only' (n = 6) (Table 2). IBI was defined as 'bacteraemia and/or bacterial meningitis' in all studies. 'SBI' most commonly referred to bacteraemia, meningitis and/or UTI, however 17 distinct 'SBI' infection combinations were identified, including one model[55] that generated predictions for three site-specific infections (pneumonia, bacteraemia or UTI) individually in lieu of a single composite outcome. The study-specific criteria **(setting-specific factors (d14)**) used to establish the **diagnosis of infection (d21)** varied with regards to culture thresholds (i.e. **bacterial growth/detection on culture (d18)**, including number, type and/or quantity of pathogen/s grown), microscopy requirements **(i.e. evidence of local inflammation on microscopy (d19)**, such as presence and/or quantity of leukocytes) and inclusion of clinical findings (e.g. **evidence of focal signs of infection (d11)**) and/or **additional investigations (d20)**, such as imaging findings (Table 3). Reported target outcome prevalence ranged between 0.3%-10.4% for bacteraemia, 1.0%-15.8% for 'IBI', and 0.8%-37.4% for 'SBI' composites (Table 2). Relationships between study cohort selection characteristics, management decision/interventions, operational factors and whether or not a child with the underlying **presence of target infection/s (d9)** receives a **diagnosis of infection (d21)** are illustrated in Fig. 4.

Table 2
Site/s of infection included in target outcome definitions and reported study prevalence by infection combination

Prediction target outcome	Site/s of bacterial infection included in prediction target outcome definitions								Studies with target outcome definition		
	Blood stream	BM	UTI	Pneumonia	BJI*	SSTI**	GI	'Other'***	No. of studies	Study outcome prevalence	Study ref.
Bacteraemia	+								6	0.3% – 10.4%	91,92 41,52,56,93
IBI	+	+							14	1.0% – 15.8%	40,43,44,48,51,53,62,63,94–99
SBI	+	+	+						7	7.1% – 22.4%	45,54,64,66–68,100
	+	+	+				+		1	5.4%	109
	+	+	+		+	+	+		1	9.9%	104
	+	+	+	+		+	+		2	8.7% – 10.1%	49,107
	+	+	+		+		+		1	4.5%	108
	+	+	+	+	+	+	+	+	5	12.6% – 28.9%	46,59,60,101,114
	+	+	+	+				+	2	23.6% – 37.4%	103,106
	+	+	+	+	+	+	+		3	15.7% – 29.0%	102,105,113
	+	+			+				1	1.1%	57
	+		+	+					1	7.1%	55
	+	+	+	+	+			+	1	23.9%	47
	+	+	+	+			+	+	3	23.2% – 26.0%	70,111,112
	+	+	+	+	+				2	3.9% – 19.8%	58,115
	+	+	+	+	+	+			1	0.8%	42
	+	+	+	+					1	26.7%	50
	+	+	+	+		+		+	1	3.8%	69
	+	+	+	+	+			+	3	8.3% – 24.1%	61,65,110

Table 3
Diagnostic criteria definitions for site-specific infections

Site of infection	Definition/s (number of studies)
Bacteraemia	Growth of pathogen in blood culture (n = 56)
Bacterial meningitis	Growth of pathogenic bacteria in CSF: only (n = 39); or focal abnormal radiological findings or consensus diagnosis (n = 4); or clinical signs and signs of pleocytosis on microscopy (n = 2); plus signs of pleocytosis on microscopy (n = 1); or positive blood culture and sterile CSF with pleocytosis +/- biochemical CSF changes (n = 1). Discharge diagnosis of 'bacterial meningitis' (n = 1)
UTI, by specimen collection method	Suprapubic aspirate (n = 12): Single urinary pathogen growth > 10^3 CFU/mL (n = 6); Any bacterial growth (n = 4); Single urinary pathogen growth >5x10^4 CFU/mL (n = 1); Not specified (n = 1) Catheter (n = 18): Single urinary pathogen growth > 10^4 CFU/mL (n = 8); Single urinary pathogen growth 1-5x10^4 CFU/mL plus leukocytes > 5 WCC/hpf (n = 5); Single urinary pathogen growth >5x10^4 CFU/mL (n = 4); Single urinary pathogen growth > 10^5 CFU/mL (n = 1); Single urinary pathogen growth > 10^3 CFU/mL (n = 1); Not specified (n = 1) Clean catch (n = 5): Single urinary pathogen growth > 10^5 CFU/mL (n = 3); Single urinary pathogen growth > 10^4 CFU/mL (n = 1); Single urinary pathogen growth >5x10^4 CFU/mL (n = 1) Bag (n = 3): Single urinary pathogen growth > 10^5 CFU/mL with clinical symptoms (n = 2); Single urinary pathogen growth > 10^5 CFU/mL in two or more samples (n = 1) Method not specified (n = 20): Pyelonephritis with Single urinary pathogen growth > 10^5 CFU/mL and abnormal DMSA (n = 4); Single urinary pathogen growth > 10^5 CFU/mL (n = 3); Single urinary pathogen growth > 10^4 CFU/mL with leuks > 5 and treated as UTI (n = 2); Single urinary pathogen growth > 10^4 CFU/mL (n = 1); Single urinary pathogen growth > 10^5 CFU/mL with clinical symptoms (n = 1); Not specified (n = 10)
Pneumonia	Infiltrate and/or consolidation on CXR/radiological signs (n = 11); Respiratory signs plus CXR findings (n = 3); Consolidation on CXR plus any of: positive blood culture, pleural fluid culture/PCR, *Mycoplasma* serology (n = 1); Presence of consolidation on CXR plus positive culture on sputum or pleural fluid (n = 1); Growth of pathogen from pleural fluid (n = 1); Not specified (n = 4)
BJI	Growth of pathogen from site (n = 4); Based on consensus diagnosis/discharge diagnosis (n = 4); Positive bacterial cultures, focal abnormal radiological findings or consensus diagnosis (n = 2); Clinical and imaging findings or positive blood culture for osteomyelitis, aspiration with growth of pathogen for septic arthritis (n = 1); Physical examination (n = 1); Not specified (n = 9)
SSTI	Consensus diagnosis (n = 4); positive bacterial cultures, focal abnormal radiological findings or consensus diagnosis (n = 3); Growth of pathogen from site (n = 2); Deep abscess on CT and surgical exploration (n = 1); Suppurative infection (n = 1); Physical examination (n = 1); Not specified (n = 1)
Bacterial gastroenteritis	Growth of pathogen (n = 12); Growth of *Salmonella, Shigella, Campylobacter* (n = 1); Growth of *Salmonella* (n = 1); Not specified (n = 5)

Abbreviations: BJI bone and joint infection; CFU colony forming units; CSF cerebrospinal fluid; CT computed tomography scan; CXR chest X-ray; DMSA dimercaptosuccinic acid scan; hpf high-powered field; PCR polymerase chain reaction; SSTI skin and soft tissue infection; UTI urinary tract infection; WCC white cell count

Model derivation and algorithm types

Model derivation methods were described in 49 studies (Supplementary material C); 12 studies reported the development of two or more models (range 2–11). Internal model validation was described in 23 studies, and external validation was described in six.[56–61] Five broad algorithm types were identified, comprising rule-based, sequential decision-tree (SDT), score-based, statistical formula (regression) and machine-learned models (Table 4). Rule-based (n = 16) and SDT models (n = 13) were the most frequent algorithm types; an increase in more complex statistical and machine-learned models was observed over the last decade (Fig. 3). All rule-based, SDT and score-based algorithms were presented within the study publication. Of the 11 statistical CPMs, seven were published as an equation, two as tabulated coefficients and four were not published. Three of the eight machine-learned model derivation studies referenced an online source to model code;[62–64] none were accessible as a tool for implementation.

Table 4
Clinical prediction model algorithm category descriptions

Algorithm type	Description
Rule-based model	A set of deterministic, predefined rules or logic that, if are met/not met, provide an estimate of the risk/probability of an outcome being present, e.g. if 'A' and 'B' and 'C' are true, then outcome 'D' is likely/unlikely).
Score-based model	Points are assigned to the presence/absence or degree of severity of predictor variables to derive a total score, with 'cut-offs' or thresholds allocated to indicate the risk/probability of an outcome being present.
Sequential Decision Tree (SDT) model	A tree-like model of decisions consisting of multiple steps of 'if then, else' logic, wherein each sequential chain of events along that 'branch' until its terminal end adds incremental value to the risk or probability an outcome being present.
Statistical formula (regression) model	A mathematical formula that represents the probability of a target outcome as a linear weighted sum of the relevant predictor variables, e.g. linear regression or logistic regression models.
Machine-learned model	An underlying computational method (e.g. deep-learning, artificial intelligence) that identifies patterns in data to make predictions based on a given set of predictor variables, e.g. support vector machine, neural network, and random forests methods.

Abbreviations: SDT sequential decision tree

Model predictor variables

Figure 5 depicts how diverse predictor variables may influence on or be influenced by an individual's **susceptibility to infection progression (d23)** or serve as reportable or observable indicators of the underlying host response (**focal/systemic inflammation (d24)**) to the **presence of target infection/s (d9)**, and how their observed values may be influenced at different time points by other predictors and/or clinical **management decisions/interventions (s5)**. Predictor variable number varied broadly between models (range 2–28), and included differing combinations of patient **background features (s1)**, previous or current **symptoms (s2)** (features of the history of presenting complaint) or **signs (s2)** (examination findings), and/or **investigation results (s6)** (Table 1; Fig. 5). The median number of predictor variables was greater for statistical and machine-learned models than for other algorithm types (median 7 [range 3–28] and 9 [range 4–25] versus median 5 [range 2–14], respectively). Investigation results and examination signs were the most frequent variable types, used in 49 and 45 models respectively; five models[50,54,65–67] used investigation result variables only, whereas seven models[42,43,55,59,68–70] used no investigation variables. The most frequently used individual variables were urinalysis (n = 30), 'appearance' (n = 26), c-reactive protein (CRP) (n = 25), white cell count (WCC) (n = 25), age (n = 24), maximum recorded temperature (n = 20) and procalcitonin (PCT) (n = 16)(Table 1).

Model prediction outputs and predictive performance

Most CPM prediction outputs were presented as categorical risk-groups (e.g. 'high-risk' or 'low-risk') intended to either 'rule-in' or 'rule-out' the target outcome (n = 39) (Table 1). Fourteen models provided a probabilistic estimate of the target outcome being present, e.g. likelihood quantified on a continuum from 0% (impossible) to 100% (certain). Two CPMs presented both estimated probability and risk-group categorisation. One study[46] did not describe its model output type. Four CPMs, all machine-learned, could predict the likelihood of the target outcome when data for one or more variable was missing (Supplementary material C).[44,58,63,64] Twenty-five models (13 rule-based, 1 score-based and 11 SDT) were able to categorise into an off-target risk-group (e.g. classification as high-risk in a low-risk 'rule-out' rule) if other available predictors had already established that the target risk-group criteria were not met.

All studies reported at least one performance characteristic, with sensitivity ranging from 5.3%-100%, specificity from 26.7%-99.2%, PPV from 1.3%-80%, NPV from 85%-100% and AUROC from 0.768-1.0 (Supplementary material C). Both the highest reported specificity (99.2%) and lowest reported sensitivity (5.3%) were observed in one 'rule-in' model for SBI.[65] All 'rule-out' models reported sensitivity > 85% and NPV > 95%, however with specificity and PPV varying broadly. Supplementary Material D illustrates using causal DAGs how differences in cohort selection and study design may alter baseline outcome prevalence, thereby contributing to variation in predictive performance and limiting direct comparison between models.

Discussion

This scoping review systematically characterises the key features of available CPMs developed to estimate the risk of serious infection in febrile children, examining how methodological and design differences may influence model performance, applicability and usability across clinical settings. Specifically, we identify the principal sources of this heterogeneity, including differences in derivation cohort characteristics and target outcome definitions, as well as substantial variation in algorithm complexity, data requirements, and approaches to handling missing data, underscoring the challenges clinicians may encounter when evaluating model suitability for their own context. By contextualising key concepts within causal DAGs, we illustrate how these features may influence case ascertainment, predictive performance, and model transportability, while providing a framework that links a CPM's intended purpose, methodological design, and clinical application, and identifies opportunities to better align core definitions and design elements with the clinical decisions the model is intended to support.

Designing models for real-world, relevant decision-making

Model derivation cohorts differed substantially in their selection criteria, including age cut-offs, temperature thresholds, and use of additional exclusions, and few included older children or those assessed in ambulatory care. While narrowly defined inclusion criteria may enhance a study's internal validity, such algorithms are unlikely to be transportable to broader, more heterogenous cohorts or settings where the pre-test disease probability is likely to differ and in whom the factors affecting this risk may also vary.[20,71,72] Mapping these characteristics within a causal framework illustrates how each additional selection

criterion may systematically alter the baseline risk profile of the derivation cohort by preferentially including or excluding children at higher or lower risk of the target outcome, which, in turn, may influence its observed prevalence and consequently the model's predictive performance. Supplementary material D provides an example of such an effect on the observed prevalence of IBI. Furthermore, almost half of all studies excluded a number of otherwise eligible participants due to missing results or incomplete assessment. As illustrated in Fig. 4, the availability of test results is contingent on the decision to perform the test, which may be influenced by clinicians' perceptions of its diagnostic utility and practical feasibility within a given clinical setting.[73] Patients with missing data therefore often differ systematically from those with complete datasets and their exclusion may introduce substantial selection bias.[23] Incorporating eligibility criteria as predictor variables, rather than as study exclusions, and employing statistical approaches to impute participants' missing data, may improve the representativeness of study derivation cohorts and improve model applicability across broader patient populations.[6,23,74]

Although many models sought to predict 'SBI', study target outcomes varied considerably with respect to both the infection groups included and the diagnostic criteria used to ascertain them, contributing to wide differences in reported prevalence and limiting meaningful comparisons of model performance. These effects are illustrated in Fig. 4, which distinguishes the true presence of infection from its diagnosis based on study-specific operational definitions, highlighting how observed outcomes may be shaped by multiple underlying causal processes and are therefore susceptible to uncertainty and misclassification. As such, highly specific criteria which include only microbiologically-confirmed infection may support a 'rule-in' diagnostic approach, but risks under-identifying less clear cut cases; conversely, highly sensitive criteria that also include clinically-suspected, but not microbiologically confirmed, cases may support a safe and effective 'rule-out' of infection, but at the expense of more false positives and unnecessary interventions.[75] Furthermore, although the use of composite outcomes is pragmatic, aggregating common infections (e.g. UTIs) with less common, but important, infections (e.g. meningitis) may also obscure a model's ability to detect the latter.[76] Clinician willingness to accept definitions that increase false negatives or false positives should be based on balancing the consequences of over-treatment or missing an infection of concern;[77] models capable of generating individual risk estimates for specific infections from a shared set of predictors may improve comparisons of tool performance and better inform more targeted clinical management. Causal frameworks such as those depicted may support a systematic approach to evaluate the diagnostic trade-offs inherent in differing target outcome definitions and to determine their acceptability to clinicians in the context of a tool's intended use.[38]

Designing clinically usable and interpretable models for real-world implementation

Beyond differences in study design and target outcomes, considerable heterogeneity was observed in the complexity of modelling approaches. Rule-based and SDT algorithms were the most commonly developed model types and were mostly straightforward deterministic algorithms with relatively few variables. While this may enable clear interpretation of model outputs and practical 'bed-side' integration, it may constrain the ability to provide probabilistic outcome estimates, utilise continuous numerical data directly, or generate predictions when input data are missing.[78] In contrast, the identified score-based and statistical models were able to incorporate a broader range of data types as predictors (i.e. continuous, binary and count) and could clarify each variable's contribution to the final risk estimate. Whereas score-based models simply added the predetermined points associated with each possible value of each predictor, most statistical models involved more complex calculations and often used greater number of variables.[78,79] The identified machine learning-based CPMs generally demonstrated strong predictive performance and several could accommodate for missing input data; however computational methods were often opaque, limiting interrogation of how outputs were derived and, unlike statistical regression models, the relative importance of each predictor could not be represented as a weight or coefficient within a published algorithm for clinical implementation.[26,78] Most model prediction outputs were presented as 'risk-group' categories that could directly inform decision-making; such outputs were commonly used in rule-based, score-based and SDT models and may be particularly desired in settings where clinician experience or access to senior support varies. However, risk thresholds were often conservative, prioritising either specificity or sensitivity at the expense of overall performance and limiting the proportion of patients without the target outcome classified as low-risk and therefore eligible for less conservative management. A number of models presented predictions as a probabilistic estimate, which can allow for more granular and individualised risk assesments; however, these were often presented without recommended thresholds for action, placing responsibility on the clinican to determine the clinical relevance of an output to inform further management. Most models utilised relatively few variables to generate predictions, however several required a large number of predictor variables (up to 28 variables), which may affect their practical implementation when time or data availability is constrained.[80] Clinician trust and engagement with CPMs have been shown to depend on the interpretability of their algorithms and prediction outputs, and their ease of integration into existing clinical workflows.[24,80,81] The continued development of less complex models may reflect a clinical preference for tools that prioritise greater interpretability and usability, over more granular risk estimatation and potentially greater predictive performance, and highlights the need to improve the explainability of both simple and more complex prediction models under the context of explicit decision-making to support their adoption in clinical practice.[81]

Model predictor variables types were predominantly selected on the basis of their statistical significance, with variable incorporation of expert input regarding their *a priori* importance, which may increase the risk of spurious associations and compromise model robustness and interpretability.[82] Predictors selected based on *a priori* understanding of their likely significance to the causal mechanisms underpinning the risk of infection can enhance the biological plausibility of data-driven approaches, increase clinician trust and strengthen the transportability of prediction tools across settings.[83,84] Mapping predictor variables within a causal framework provided insight into their interdependencies and the latent processes they represent, illustrating how some variable types have a more intuitive conceptual relationship with the presence of infection than others or may be more readily available in practice, e.g. those relating to patient history, signs and symptoms. However, the interpretation of signs and symptoms may be subjective, potentially introducing variability in predictor data quality. Investigation results can be obtained more objectively, however some tests may not be available or routinely used in some clinical settings[76] or may only become available later in the assessment process.[80] Several studies addressed this issue by completely avoiding the use of biomarkers as predictors, or adapted their algorithms to use more commonly available investigations (e.g. CRP) in place of others (e.g. PCT). However, most models were unable to generate predictions when input data were unavailable, potentially limiting their real-world usability and highlighting the need for future algorithms that can accommodate missing or sequentially accruing data.[28] It is noteworthy that only one of the identified models[67] incorporated 'viral test results' as an

predictor variable and none used temporal factors, such as 'season'. Viruses are the major differential diagnosis for a child presenting with fever and many have highly seasonal transmission, which may affect the predictive performance of tools over time;[3,85] incorporating patient-specific viral test results and local pathogen surveillance data as predictors may enhance model performance across diverse geographical and temporal settings.[86,87] However, it is important to note that integrating these factors within deterministic (rule-based) models may be challenging as their influence on SBI risk is probabilistic, i.e. a positive viral result may reduce the likelihood of SBI but does not exclude it.[88] Model algorithms based on causal probabilistic frameworks may enable more effective use of such predictors to support more individualised clinical decision-making.[89,90]

Future directions

By characterising key study and model features within a causal DAG framework, this scoping review provides a structured approach for evaluating the clinical applicability and transportability of available CPMs, supporting clinicians in model selection and highlighting opportunities for methodological improvements. Incorporating causal DAGs *a priori* into future model development may support the alignment of disease definitions and guide the selection of predictors that capture stable causal relationships, thereby enhancing model transportability across clinical settings. Furthermore, quantifying the strengths of these causal relationships may provide a basis for explainable prediction models that can integrate new evidence (e.g. patient symptoms) with prior information (e.g. baseline prevalence) while accommodating missing variables, thereby improving predictive performance and enhancing clinical utility.[77] Importantly, further work is needed to understand end-user preferences for CPM interface design, including output presentation format, the optimal balance between model complexity and usability, and methods for interrogating underlying model mechanisms. Addressing these factors will be critical to developing fit-for-purpose tools that are trusted and adopted in routine clinical practice.

Strengths and limitations

A key strength of this review is its systematic approach to study identification and data synthesis, including the use of causal DAGs, which enabled comprehensive characterisation of the principal design and development features of more than 50 original CPMs published over the past four decades. However, not all relevant studies may have been identified through the selected databases and search strategy. Furthermore, because only original model development studies were included, subsequent evaluations of model performance, validation, and clinical impact were not captured. We constructed causal DAGs incorporating both identified study characteristics and additional observable and latent variables deemed important to capture the principal causal processes underlying the clinical domain. However, as DAGs are inherently simplified representations of real-world processes, they are unlikely to capture all factors contributing to the complex causal pathways underlying fever and infection and alternative causal structures may also provide equally valid representations of these processes. As such, the presented DAGs should be viewed as a starting point for further refinement and discussion. Lastly, studies pertaining to specific sub-populations (e.g. immunocompromised patients) or non-diagnostic predictions (e.g. prognostic models) were not included, and a formal risk of bias assessment was not performed for each individual study. However, by mapping key concepts to causal DAGs, we expect that these findings relating to model applicability and generalisability are translatable and can inform the appraisal and future development of CPMs intended for other cohorts and conditions.

Conclusion

Available CPMs developed to predict SBI or IBI in febrile children demonstrate substantial heterogeneity, extending well beyond predictive performance alone. By systematically characterising these differences, this review provides a practical reference to support clinicians in evaluating the suitability of available models for their clinical setting and to guide researchers in addressing these methodological limitations. We demonstrate how causal DAGs can provide a transparent framework for evaluating model transportability and clinical applicability; their incorporation into future model development may help align core design elements and support the development of clinically relevant tools that better bridge the gap between research and practice.

Abbreviations

AUROC area under the receiver operating characteristic; BJI bone and joint infection; CPM clinical prediction model; CRP c-reactive protein; DAG directed acyclic graph; ED emergency department; IBI invasive bacterial infection; NPV negative predictive value; PCT procalcitonin; PPV positive predictive value; SBI serious bacterial infection; SDT sequential decision-tree; SSTI skin and soft tissue infection; UTI urinary tract infection; WCC white cell count

Declarations

Ethics approval and consent to participate: Ethics approval and consent to participate was not required

Consent for publication: Not applicable

Availability of data and materials: Not applicable

Competing interests/Conflict of Interest Disclosures (includes financial disclosures): The authors have no conflicts of interest relevant to this article to disclose.

Funding/Support: Author AOM was supported by a Western Australia Health Translation Network Early Career Fellowship, the Australian Government's Medical Research Future Fund (MRFF) as part of the Rapid Applied Research Translation Project, a National Health and Medical Research Council (NHMRC) Postgraduate Scholarship (1191465), a Wesfarmers Centre of Vaccines & Infectious Diseases Ad Hoc Postgraduate Top Up Scholarship and an Australian Government Research Training Program Fees Offset. Author TS is supported by a NHMRC Career Development Fellowship (GNT1111657).

Role of Funder/Sponsor (if any): The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Authors' contributions:

Dr Ariel Mace conceptualised and designed the study, designed the data collection instrument, extracted, analysed and interpreted data, drafted the initial manuscript, and critically reviewed and revised the manuscript.

Dr Andrew Martin conceptualised and designed the study, designed the data collection instruments, and critically reviewed and revised the manuscript.

Professor Peter Richmond conceptualised and designed the study, supervised data analyses and interpretation of data, and critically reviewed and revised the manuscript.

Professor Thomas Snelling conceptualised and designed the study, supervised data analyses and interpretation of data, and critically reviewed and revised the manuscript.

Dr Yue Wu conceptualised and designed the study, substantially contributed to data extraction, analyses and interpretation, and critically reviewed and revised the manuscript for important intellectual content.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Acknowledgements: The authors acknowledge the contributions of the Wesfarmers Infectious Diseases community reference group for their review, feedback and endorsement of the broader research program that has informed this scoping review.

References

1. Sands R, Shanmugavadivel D, Stephenson T, Wood D. Medical problems presenting to paediatric emergency departments: 10 years on. Emerg Med J. 2012;29(5):379–82.
2. Pantell RH, Naber M, Lamar R, Dias JK. Fever in the first six months of life: risks of underlying serious infection. Clin Pediatr. 1980;19(2):77–82.
3. DePorre AG, Aronson PL, McCulloh RJ. Facing the ongoing challenge of the febrile young infant. Crit Care. 2017;21:1–8.
4. Schlapbach LJ, Straney L, Alexander J, et al. Mortality related to invasive infections, sepsis, and septic shock in critically ill children in Australia and New Zealand, 2002–13: a multicentre retrospective cohort study. Lancet Infect Dis. 2015;15(1):46–54.
5. Brennan CA, Somerset M, Granier SK, Fahey TP, Heyderman RS. Management of diagnostic uncertainty in children with possible meningitis: a qualitative study. Br J Gen Pract. 2003;53(493):626.
6. Thompson M, Van den Bruel A, Verbakel J, et al. Systematic review and validation of prediction rules for identifying children with serious infections in emergency departments and urgent-access primary care. Health Technol Assess. 2012;16(15):1–100.
7. Díaz MG, García RP, Gamero DB, et al. Lack of accuracy of biomarkers and physical examination to detect bacterial infection in febrile infants. Pediatr Emerg Care. 2016;32(10):664–8.
8. Leigh S, Grant A, Murray N, et al. The cost of diagnostic uncertainty: a prospective economic analysis of febrile children attending an NHS emergency department. BMC Med. 2019;17(1). 10.1186/s12916-019-1275-z.
9. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature. JAMA. 1993;270:2093–5.
10. Wolff RF, Moons KG, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–8.
11. Irwin A, Wickenden J, Le Doare K, Ladhani S, Sharland M. Supporting decisions to increase the safe discharge of children with febrile illness from the emergency department: a systematic review and meta-analysis. Arch Dis Child. 2016;101(3):259–66.
12. Huppler AR, Eickhoff JC, Wald ER. Performance of low-risk criteria in the evaluation of young infants with fever: review of the literature. Pediatrics. 2010;125(2):228–33.
13. Umana E, Norman-Bruce H, Waterfield T. Scoping review of clinical decision aids in the assessment and management of febrile infants under 90 days of age. BMC Pediatr. 2025;25(1):274.
14. Foster LZ, Beiner J, Duh-Leong C, et al. Implementation of febrile infant management guidelines reduces hospitalization. Pediatr Qual Saf. 2020;5(1):e252.
15. Dingle E, Pelletier JH, Forbes ML, Rajbhandari P. Resource utilization and cost in management of febrile infants after the 2021 clinical guideline. Pediatrics. 2025;155(2):e2024068028.
16. Mercurio L, Hill R, Duffy S, Zonfrillo MR. Clinical practice guideline reduces evaluation and treatment for febrile infants 0 to 56 days of age. Clin Pediatr. 2020;59(9–10):893–901.
17. Roukema J, Steyerberg EW, Van der Lei J, Moll HA. Randomized trial of a clinical decision support system: impact on the management of children with fever without apparent source. J Am Med Inform Assoc. 2008;15(1):107–13.
18. Meehan WP III, Fleegler E, Bachur RG. Adherence to guidelines for managing the well-appearing febrile infant: assessment using a case-based, interactive survey. Pediatr Emerg Care. 2010;26(12):875–80.

19. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson S-A. Computerized clinical decision support for prescribing: provision does not guarantee uptake. J Am Med Inform Assoc. 2010;17(1):25–33.
20. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med Jan. 2015;6(1):W1–73. 10.7326/M14-0698.
21. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med. 2006;144(3):201–9.
22. Toll D, Janssen K, Vergouwe Y, Moons K. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol. 2008;61(11):1085–94.
23. Oostenbrink R, Thompson M, Steyerberg EW. Barriers to translating diagnostic research in febrilechildren to clinical practice: a systematic review. Arch Dis Child. 2012;97(7):667–72.
24. Evans RP, Bryant LD, Russell G, Absolom K. Trust and acceptability of data-driven clinical recommendations in everyday practice: A scoping review. Int J Med Informatics. 2024;183doi. 10.1016/j.ijmedinf.2024.105342.
25. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206–15.
26. Knop M, Weber S, Mueller M, Niehaves B. Human factors and technological characteristics influencing the interaction of medical professionals with artificial intelligence–enabled clinical decision support systems: literature review. JMIR Hum Factors. 2022;9(1):e28639.
27. Moher D, Glasziou P, Chalmers I, et al. Increasing value and reducing waste in biomedical research: who's listening? Lancet. 2016;387(10027):1573–86.
28. Cowley LE, Farewell DM, Maguire S, Kemp AM. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. Diagn Prognostic Res. 2019;2019/08/22(1):16. 10.1186/s41512-019-0060-y.
29. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009;374(9683):86–9. 10.1016/S0140-6736(09)60329-9.
30. van Maaren MC, Hueting TA, Völkel V, van Hezewijk M, Strobbe LJ, Siesling S. The use and misuse of risk prediction tools for clinical decision-making. Breast Jun. 2023;69:428–30. 10.1016/j.breast.2023.01.006.
31. Van den Bruel A, Thompson MJ, Haj-Hassan T et al. Diagnostic value of laboratory tests in identifying serious infections in febrile children: systematic review. BMJ. 2011;342.
32. Van den Bruel A, Haj-Hassan T, Thompson M, Buntinx F, Mant D. Diagnostic value of clinical features at presentation to identify serious infection in children in developed countries: a systematic review. Lancet. 2010;375(9717):834–45.
33. Collins GS, Moons KG. Comparing risk prediction models. BMJ May 24. 2012;344:e3186. 10.1136/bmj.e3186.
34. Tennant PW, Murray EJ, Arnold KF, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. Int J Epidemiol. 2021;50(2):620–32.
35. Williams TC, Bach CC, Matthiesen NB, Henriksen TB, Gagliardi L. Directed acyclic graphs: a tool for causal studies in paediatrics. Pediatr Res. 2018;84(4):487–93.
36. Pearl J. Models, reasoning and inference. Camb UK: CambridgeUniversityPress. 2000;19(2):3.
37. Fehr J, Piccininni M, Kurth T, Konigorski S. Assessing the transportability of clinical prediction models for cognitive impairment using causal models. BMC Med Res Methodol. 2023;23(1):187.
38. Piccininni M, Konigorski S, Rohmann JL, Kurth T. Directed acyclic graphs and causal thinking in clinical risk prediction modeling. BMC Med Res Methodol Jul. 2020;2(1):179. 10.1186/s12874-020-01058-z.
39. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: scoping reviews. JBI Man Evid synthesis. 2020;169(7):467–73.
40. Berkowitz CD, Uchiyama N, Tully SB, et al. Fever in infants less than two months of age: spectrum of disease and predictors of outcome. Pediatr Emerg Care. 1985;1(3):128–35.
41. Broner CW, Polk SA, Sherman JM. Febrile infants less than eight weeks old: predictors of infection. Clin Pediatr. 1990;29(8):438–43.
42. Van den Bruel A, Aertgeerts B, Bruyninckx R, Aerts M, Buntinx F. Signs and symptoms for diagnosis of serious infections in children: a prospective study in primary care. Br J Gen Pract. 2007;57(540):538.
43. Pantell RH, Newman TB, Bernzweig J, et al. Management and outcomes of care of fever in early infancy. JAMA. 2004;291(10):1203–12.
44. Ballard DW, Huang J, Sharp AL, et al. An all-inclusive model for predicting invasive bacterial infection in febrile infants age 7–60 days. Pediatr Res. 2024;96(3):759–65.
45. Chong S-L, Niu C, Ong GY-K, et al. Febrile infants risk score at triage (FIRST) for the early identification of serious bacterial infections. Sci Rep. 2023;13(1):15845.
46. Chang SSY, Lim AZ, Ong GY-K, et al. Predictors of serious bacterial infections using serum biomarkers in an infant population aged 0 to 90 days: a prospective cohort study. BMJ paediatrics open. 2021;5(1):e000861.
47. Berger R, Berger M, van Steensel-Moll H, Dzoljic-Danilovic G, Derksen-Lubsen G. A predictive model to estimate the risk of serious bacterial infections in febrile infants. Eur J Pediatrics. 1996;155:468–73.
48. Aronson PL, Shabanova V, Shapiro ED, et al. A prediction model to identify febrile infants ≤ 60 days at low risk of invasive bacterial infection. Pediatrics. 2019;144(1):e20183604.
49. Herr SM, Wald ER, Pitetti RD, Choi SS. Enhanced urinalysis improves identification of febrile infants ages 60 days and younger at low risk for serious bacterial illness. Pediatrics. 2001;108(4):866–71.

50. Lacour AG, Zamora SA, Gervaix A. A score identifying serious bacterial infections in children with fever without source. Pediatr Infect Dis J. 2008;27(7):654–6.
51. Mintegi S, Bressan S, Gomez B, et al. Accuracy of a sequential approach to identify young febrile infants at low risk for invasive bacterial infection. Emerg Med J. 2014;31(e1):e19–24.
52. Crocker PJ, Quick G, McCombs W. Occult bacteremia in the emergency department: diagnostic criteria for the young febrile child. Ann Emerg Med. 1985;14(12):1172–7.
53. Yin Z, Chen Y, Zhong W, et al. A novel algorithm with paired predictive indexes to stratify the risk levels of neonates with invasive bacterial infections: A multicenter cohort study. Pediatr Infect Dis J. 2022;41(4):e149–55.
54. Kuppermann N, Dayan PS, Levine DA, et al. A clinical prediction rule to identify febrile infants 60 days and younger at low risk for serious bacterial infections. JAMA Pediatr. 2019;173(4):342–51.
55. Craig JC, Williams GJ, Jones M et al. The accuracy of clinical symptoms and signs for the diagnosis of serious bacterial infection in young febrile children: prospective cohort study of 15 781 febrile illnesses. BMJ. 2010;340.
56. Isaacman DJ, Shults J, Gross TK, Davis PH, Harper M. Predictors of bacteremia in febrile children 3 to 36 months of age. Pediatrics. 2000;106(5):977–82.
57. Hagedoorn NN, Borensztajn D, Nijman RG, et al. Development and validation of a prediction model for invasive bacterial infections in febrile children at European Emergency Departments: MOFICHE, a prospective observational study. Arch Dis Child. 2021;106(7):641–7.
58. Lee B, Chung HJ, Kang HM, Kim DK, Kwak YH. Development and validation of machine learning-driven prediction model for serious bacterial infection among febrile children in emergency departments. PLoS ONE. 2022;17(3):e0265500.
59. Urbane UN, Petrosina E, Zavadska D, Pavare J. Integrating Clinical Signs at Presentation and Clinician's Non-analytical Reasoning in Prediction Models for Serious Bacterial Infection in Febrile Children Presenting to Emergency Department. Front Pead. 2022;10doi. 10.3389/fped.2022.786795.
60. Nijman RG, Vergouwe Y, Thompson M et al. Clinical prediction model to aid emergency doctors managing febrile children at risk of serious bacterial infections: diagnostic study. BMJ. 2013;346.
61. Leroy S, Bressan S, Lacroix L, et al. Refined lab-score, a risk score predicting serious bacterial infection in febrile children less than 3 years of age. Pediatr Infect Dis J. 2018;37(5):387–93.
62. Yang Y, Wang Y-M, Lin C-HR, et al. Explainable deep learning model to predict invasive bacterial infection in febrile young infants: a retrospective study. Int J Med Informatics. 2023;172:105007.
63. Yaeger JP, Jones J, Ertefaie A, Caserta MT, Fiscella KA. Derivation of a clinical-based model to detect invasive bacterial infections in febrile infants. J Hosp Med. 2022;17(11):893–900.
64. Yaeger JP, Jones J, Ertefaie A, Caserta MT, van Wijngaarden E, Fiscella K. Using clinical history factors to identify bacterial infections in young febrile infants. J Pediatr. 2021;232:192–9. e2.
65. Vorwerk C, Manias K, Davies F, Coats TJ. Prediction of severe bacterial infection in children with an emergency department diagnosis of infection. Emerg Med J. 2011;28(11):948–51.
66. Ramgopal S, Horvat CM, Yanamala N, Alpern ER. Machine learning to predict serious bacterial infections in young febrile infants. Pediatrics. 2020;146(3).
67. Poirier C, Filion-Ouellet E, Anderson G, Burstein B. 107 Addition of CRP and Respiratory Viral Testing for Risk-Stratification of Febrile Young Infants. Paediatr Child Health. 2021;26(Supplement1):e75–6.
68. Bonadio WA, Hennes H, Smith D, et al. Reliability of observation variables in distinguishing infectious outcome of febrile young infants. Pediatr Infect Dis J. 1993;12(2):111–4.
69. Brent AJ, Lakhanpaul M, Thompson M, et al. Risk score to stratify children with suspected serious bacterial infection: observational cohort study. Arch Dis Child. 2011;96(4):361–7.
70. Roukema J, van Loenhout RB, Steyerberg EW, Moons KG, Bleeker SE, Moll HA. Polytomous regression did not outperform dichotomous logistic regression in diagnosing serious bacterial infections in febrile children. J Clin Epidemiol. 2008;61(2):135–41.
71. Ferguson L. External validity, generalizability, and knowledge utilization. J Nurs Scholarsh. 2004;36(1):16–22.
72. Gill CJ, Sabin L, Schmid CH. Why clinicians are natural bayesians. BMJ. 2005;330(7499):1080–3.
73. Dewez JE, Pembrey L, Nijman RG, et al. Availability and use of rapid diagnostic tests for the management of acute childhood infections in Europe: A cross-sectional survey of paediatricians. PLoS ONE. 2022;17(12):e0275336.
74. Heymans MW, Twisk JW. Handling missing data in clinical research. J Clin Epidemiol. 2022;151:185–8.
75. Reitsma JB, Rutjes AWS, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. J Clin Epidemiol. 2009;62(8):797–806. 10.1016/j.jclinepi.2009.02.005.
76. Pantell RH, Roberts KB, Adams WG et al. Clinical practice guideline: evaluation and management of well-appearing febrile infants 8 to 60 days old. Pediatrics. 2021;148(2).
77. Wu Y, Mascaro S, Bhuiyan M, et al. Predicting the causative pathogen among children with pneumonia using a causal Bayesian network. PLoS Comput Biol. 2023;19(3):e1010967.
78. Harper PR. A review and comparison of classification algorithms for medical decision making. Health Policy. 2005;71(3):315–31.
79. Arora P, Boyne D, Slater JJ, Gupta A, Brenner DR, Druzdzel MJ. Bayesian networks for risk prediction using real-world data: a tool for precision medicine. Value Health. 2019;22(4):439–45.

80. Kennedy G, Gallego B. Clinical prediction rules: a systematic review of healthcare provider opinions and preferences. Int J Med Informatics. 2019;123:1–10.
81. Chan SL, Lee JW, Ong MEH, et al. Implementation of Prediction Models in the Emergency Department from an Implementation Science Perspective—Determinants, Outcomes,and Real-World Impact: A Scoping Review. *Annals of Emergency Medicine*. 2023;82(1):22–36. doi:10.1016/j.annemergmed.2023.02.001.
82. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361–87.
83. Meehan AJ, Danese A. Progress and barriers to the implementation of prediction modelling in child and adolescent mental health-A commentary on Senior. () JCPP Adv Dec. 2021;1(4):e12052. 10.1002/jcv2.12052.
84. Stolper E, Van de Wiel M, Van Royen P, Van Bokhoven M, Van der Weijden T, Dinant GJ. Gut feelings as a third track in general practitioners' diagnostic reasoning. J Gen Intern Med. 2011;26(2):197–203.
85. Greenhow TL, Hung Y-Y, Herz AM, Losada E, Pantell RH. The changing epidemiology of serious bacterial infections in young infants. Pediatr Infect Dis J. 2014;33(6):595–9.
86. Fine AM, Nigrovic LE, Reis BY, Cook EF, Mandl KD. Linking surveillance to action: incorporation of real-time regional data into a medical decision rule. J Am Med Inform Assoc. 2007;14(2):206–11.
87. Rybak A, Aupiais C, Cotillon M, et al. Reassessing the Performance of the Step-By-Step Approach to Febrile Infants 90 Days of Age and Younger in the Context of the COVID-19 Pandemic: A Multicentric Retrospective Study. Pediatr Infect Dis J Sep. 2022;1(9):e365–8. 10.1097/INF.0000000000003614.
88. Perez-Porra S, Granda E, Benito H, Roland D, Gomez B, Velasco R. Prevalence of invasive bacterial infection in febrile infants =90 days with a COVID-19 positive test: a systematic review and meta-analysis</at. Emerg Med J Mar. 2024;21(4):228–35. 10.1136/emermed-2023-213483.
89. Coulthard MG. Quantifying how tests reduce diagnostic uncertainty. Arch Dis Child. 2007;92(5):404–8.
90. Ramsay JA, Mascaro S, Campbell AJ, et al. Urinary tract infections in children: building a causal model-based decision support tool for diagnosis with domain knowledge and prospective data. BMC Med Res Methodol. 2022;22(1). 10.1186/s12874-022-01695-6.
91. Crain EF, Shelov SP. Febrile infants: predictors of bacteremia. J Pediatr. 1982;101(5):686–9.
92. Gómez B, Mintegi S, Benito J, Egireun A, Garcia D, Astobiza E. Blood culture and bacteremia predictors in infants less than three months of age with fever without source. Pediatr Infect Dis J. 2010;29(1):43–7.
93. Tsai C-M, Lin C-HR, Zhang H, et al. Using machine learning to predict bacteremia in febrile children presented to the emergency department. Diagnostics. 2020;10(5):307.
94. Velasco R, Benito H, Mozún R, Trujillo JE, Merino PA. Network GftSoFIotR-S. Febrile young infants with altered urinalysis at low risk for invasive bacterial infection. a Spanish Pediatric Emergency Research Network's Study. Pediatr Infect Dis J. 2015;34(1):17–21.
95. Nguyen TH, Young BR, Poggel LE, Alabaster A, Greenhow TL. Roseville protocol for the management of febrile infants 7–60 days. Hosp Pediatr. 2021;11(1):52–60.
96. Burstein B, Alathari N, Papenburg J. Guideline-based risk stratification for febrile young infants without procalcitonin measurement. Pediatrics. 2022;149(6):e2021056028.
97. Burstein B, Wolek C, Poirier C, et al. Optimizing Management of Febrile Young Infants Without Serum Procalcitonin. Pediatrics. 2025;155(2):e2024068200.
98. Dorney K, Neuman MI, Harper MB, Bachur RG. Boston febrile infant algorithm 2.0: improving care of the febrile infant 1–2 months of age. Pediatr Qual Saf. 2022;7(6):e616.
99. Feildel-Fournial C, Danjou L, Hubert-Dibon G, et al. A simple clinical decision rule may help to rule out severe bacterial infection and spare antibiotics in febrile young children. Acta Paediatr. 2023;112(4):837–45.
100. Bachur RG, Harper MB. Predictive model for serious bacterial infections among infants younger than 3 months of age. Pediatrics. 2001;108(2):311–6.
101. Marom R, Sakran W, Antonelli J, et al. Quick identification of febrile neonates with low risk for serious bacterial infection: an observational study. Archives Disease Childhood-Fetal Neonatal Ed. 2007;92(1):F15–8.
102. Li S-Y, Yin C-H, Chen J-S, et al. A nomogram for predicting the development of serious bacterial infections in febrile term neonates: A single medical center experience in Southern Taiwan. Pediatr Neonatology. 2022;63(6):605–12.
103. Vujevic M, Benzon B, Markic J. New prediction model for diagnosis of bacterial infection in febrile infants younger than 90 days. Turk J Pediatr. 2017;59(3):261–8.
104. Dagan R, Powell KR, Hall CB, Menegus MA. Identification of infants unlikely to have serious bacterial infection although hospitalized for suspected sepsis. J Pediatr. 1985;107(6):855–60.
105. Pinto EV, Sánchez-Bayle M. Construction of a diagnostic prediction model of severe bacterial infection in febrile infants under 3 months old. Anales de Pediatría (English Edition). 2017;87(6):330–6.
106. Olaciregui I, Hernández U, Muñoz J, Emparanza JI, Landa JJ. Markers that predict serious bacterial infection in infants under 3 months of age presenting with fever of unknown origin. Arch Dis Child. 2009;94(7):501–5.
107. Baker MD, Bell LM, Avner JR. Outpatient management without antibiotics of fever in selected infants. N Engl J Med. 1993;329(20):1437–41.
108. Bonadio WA, Hagen E, Rucka J, Shallow K, Stommel P, Smith D. Efficacy of a protocol to distinguish risk of serious bacterial infection in the outpatient evaluation of febrile young infants. Clin Pediatr. 1993;32(7):401–4.

109. Baskin MN, O'Rourke EJ, Fleisher GR. Outpatient treatment of febrile infants 28 to 89 days of age with intramuscular administration of ceftriaxone. J Pediatr. 1992;120(1):22–7.
110. Irwin AD, Grant A, Williams R et al. Predicting risk of serious bacterial infections in febrile children in the emergency department. Pediatrics. 2017;140(2).
111. Bleeker S, Moons K, Derksen-Lubsen G, Grobbee D, Moll H. Predicting serious bacterial infection in young children with fever without apparent source. Acta Paediatr. 2001;90(11):1226–31.
112. Bleeker S, Derksen-Lubsen G, Grobbee D, Donders A, Moons K, Moll H. Validating and updating a prediction rule for serious bacterial infection in patients with fever without source. Acta Paediatr. 2007;96(1):100–4.
113. Nijman RG, Moll HA, Smit FJ, et al. C-reactive protein, procalcitonin and the lab-score for detecting serious bacterial infections in febrile children at the emergency department: a prospective observational study. Pediatr Infect Dis J. 2014;33(11):e273–9.
114. Nijman RG, Vergouwe Y, Moll HA, et al. Validation of the Feverkidstool and procalcitonin for detecting serious bacterial infections in febrile children. Pediatr Res. 2018;83(2):466–76.
115. Goh P, Lee S, Wong E. Predictors of serious bacterial infection in children aged 3 to 36 months with fever without source. Singapore Med J. 2006;47(4):276.

Figures

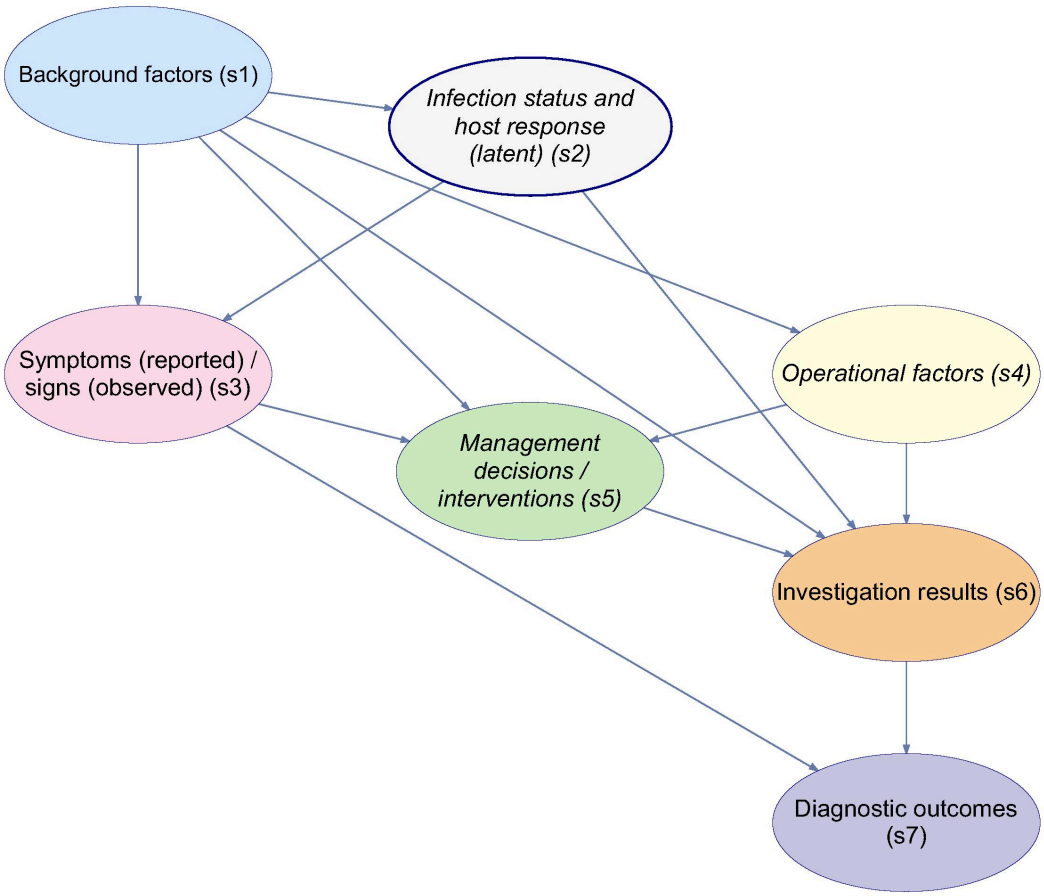


Figure 1

High-level causal DAG for diagnosing serious infections in children.

Caption: This DAG illustrates at a high-level how key patient, disease and procedural concepts mechanistically influence (positively or negatively) the presence and diagnosis of serious infections in children. The depicted DAG differentiates **diagnostic outcomes**(purple) from the underlying latent **infection status and host response** (white), which cannot be directly observed but must be inferred from reported or observable evidence or data. It also illustrates how this observable evidence, e.g. **symptoms and signs** (pink) or **investigation results** (orange), may be influenced by multiple underlying causal processes, such as **background factors** (blue), **management decisions/interventions** (green) and/or **operational factors** (yellow), that may be known or unknown.

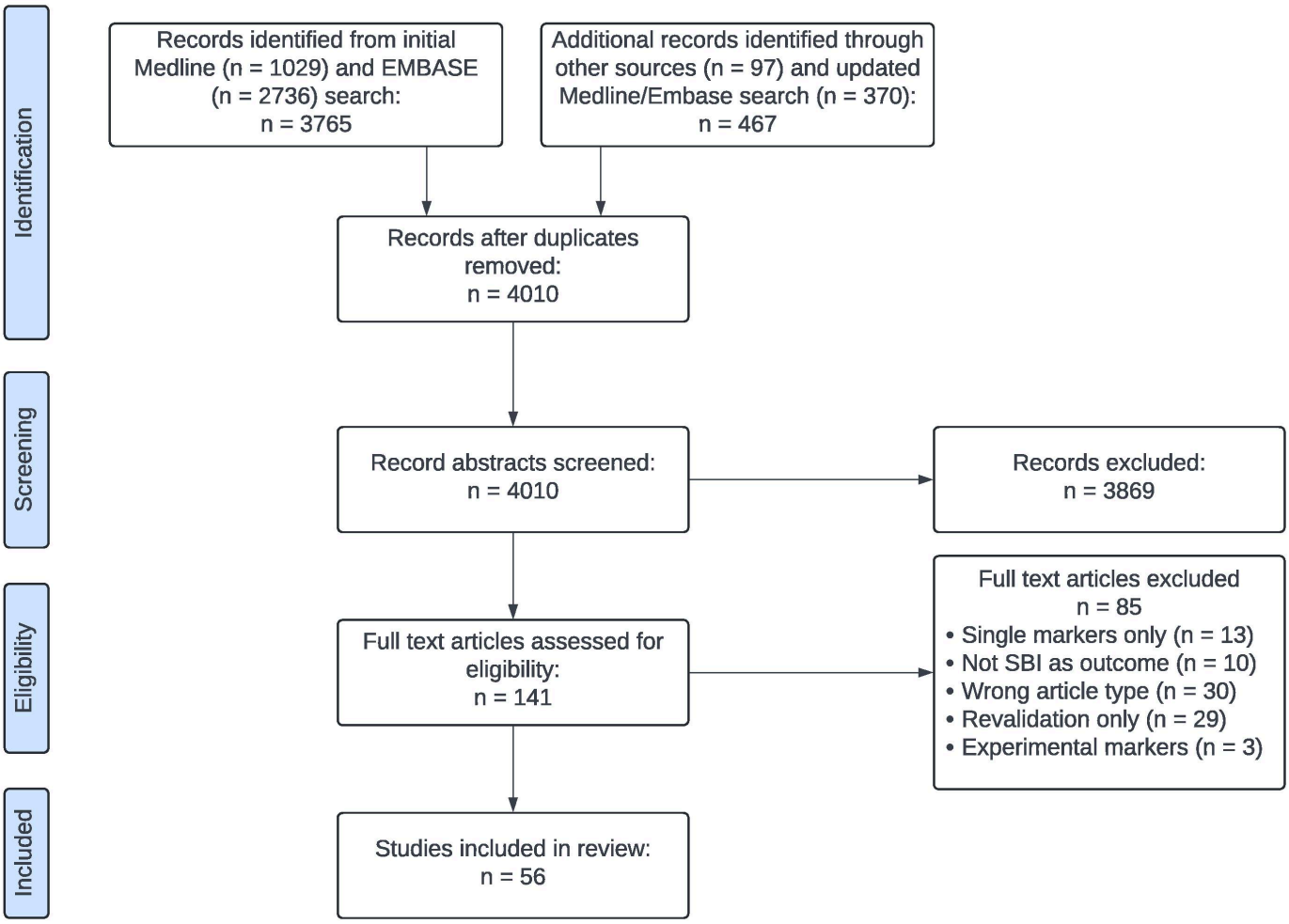


Figure 2

Study identification flow chart

Abbreviations: SBI serious bacterial infection

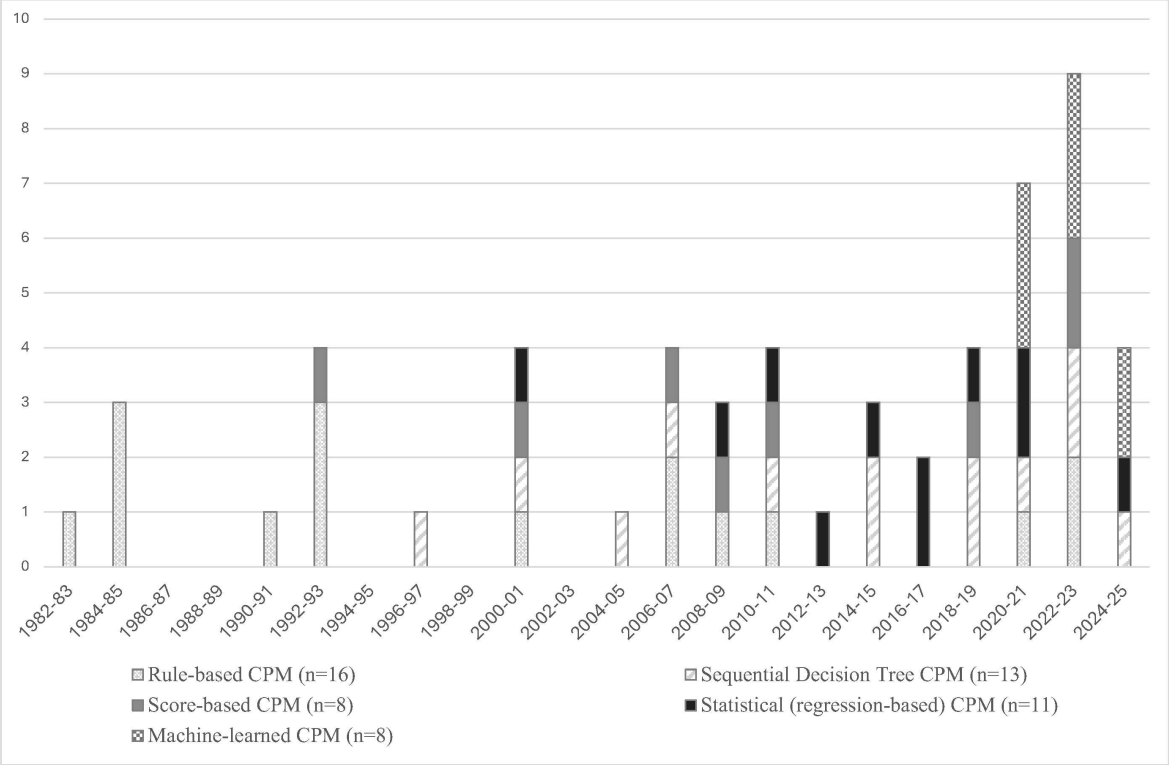


Figure 3

Study publication year by clinical prediction model algorithm type

Abbreviations: CPM clinical prediction model

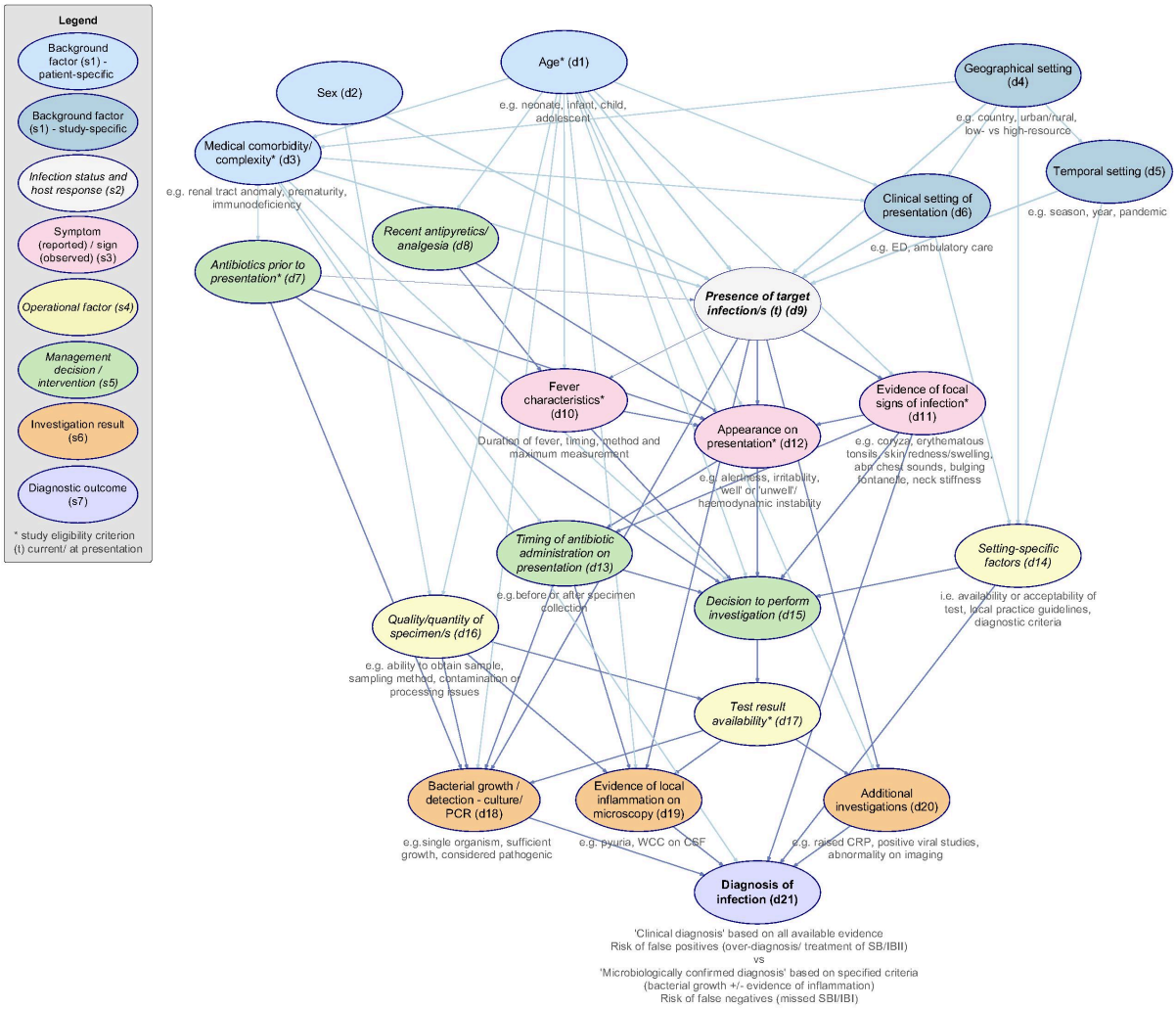


Figure 4

Causal DAG for study design, cohort selection and target outcome(s).

Caption: This DAG illustrates hypothesised causal relationships between study characteristics, including **geographical**, **temporal**, and **clinical setting**, and the likelihood of a child having the **target infection**at presentation. It demonstrates how patient background characteristics (e.g. **age, sex**, **medical comorbidities** or **complexity**) and presenting features (e.g. **appearance** or **focal signs of infection**) may influence both clinical management before presentation (e.g. **prior antibiotic** or **antipyretic/analgesic use**) and during assessment (e.g. **decisions to perform investigations** or **administer antibiotics**), as well as the subsequent results of those investigations (e.g. leukocyte counts and bacterial growth on **microscopy** or **culture**). The DAG further illustrates how **operational factors** (e.g. local access to testing or ability to obtain an adequate specimen) may be influenced by background factors (such as **clinical setting** or patient **age**), and how these interacting factors may determine whether a child satisfies the study's **operational diagnostic criteria** for the target

infection(s), thereby influencing the reported prevalence of infection within the derivation cohort. It also depicts how clinicians may assign a diagnosis of infection based on the overall clinical picture, including **symptoms, signs**, and supportive **investigation findings**, even when microbiological diagnostic criteria are not fulfilled. To note: due to the number of arrows emanating from background factors, a lighter shade of blue has been used to make the graph easier to read.

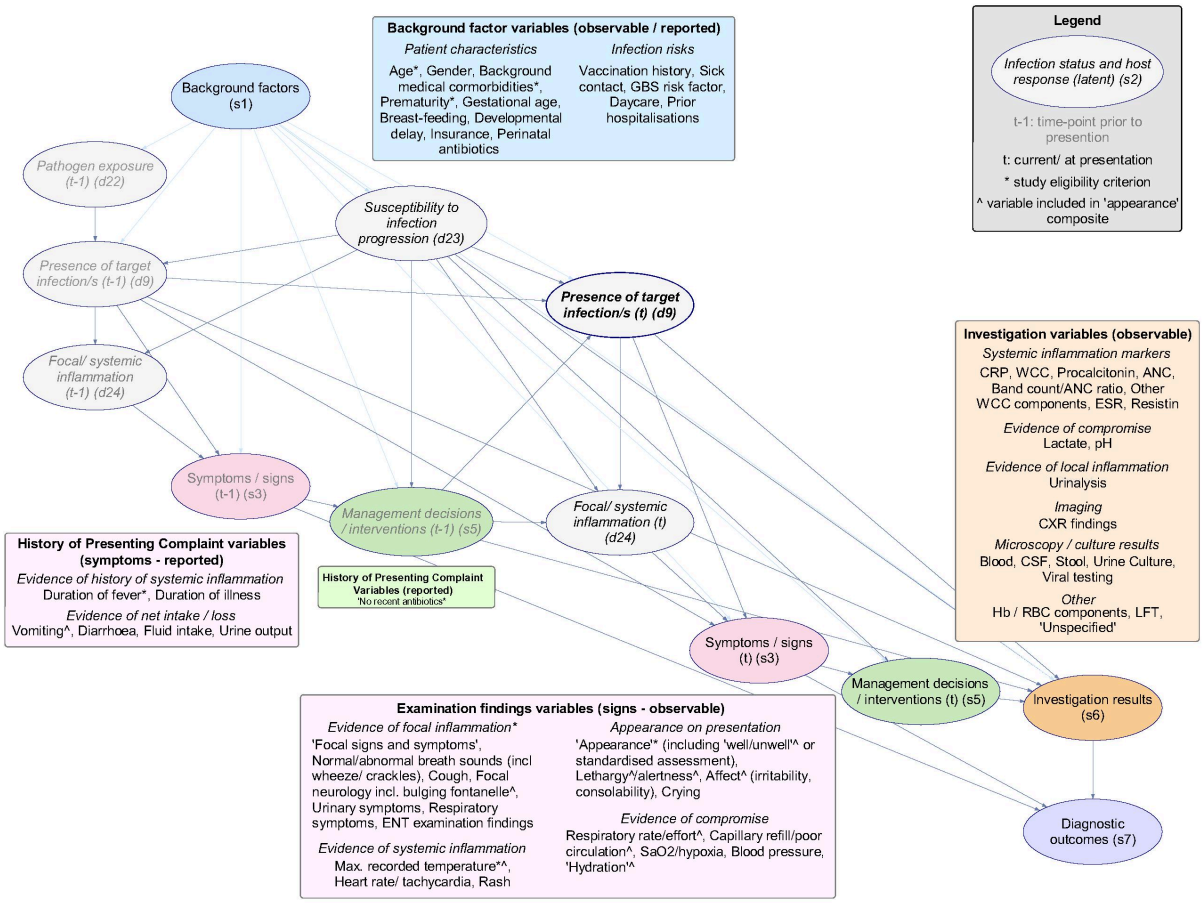


Figure 5

Causal DAG for CPM predictor variables

Caption: This DAG depicts the pathophysiological processes linking **pathogen exposure**, **host susceptibility**, and **local and systemic inflammation** to the measurable clinical features (**symptoms/signs)** and **investigation results** that may serve as predictor variables for the **presence of target infection/s** in

clinical prediction models. It illustrates that the true presence of site-specific infection and its underlying biological processes are latent (unobservable) states that can only be inferred from observable clinical evidence, including symptoms, signs, and laboratory investigations. From a causal perspective, these latent mechanisms give rise to observable features through multiple interacting pathways that may also be influenced by **background factors** and **management decisions/interventions**, such that individual predictors are rarely caused by a single process or are mutually exclusive, e.g. respiratory rate may be influenced by age, fever, pain, and pulmonary inflammation associated with respiratory infection. The DAG also illustrates temporal dependencies between predictors, whereby earlier symptoms or signs, and associated **management decisions/interventions**, may influence subsequent observations; for example, vomiting and diarrhoea at time *t–1* may contribute to dehydration and the later development of lethargy at time *t*, whereas the administration of antipyretics/analgeia or antibiotics may influence a child's apperance and their investigation results. To note: due to the number of arrows emanating from background factors, a lighter shade of blue has been used to make the graph easier to read.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MaceetalCPMscopingreviewtables.docx
- MaceetalCPMscopingreviewsuppmatABC.docx
- MaceetalCPMscopingreviewsuppmatD.pdf