Predictors of disease severity in children presenting from the community with febrile illnesses: a systematic review of prognostic studies

Arjun Chandna, Rainer Tan, Michael Carter, Ann Van Den Bruel, Jan Verbakel, Constantinos Koshiaris, Nahya Salim, Yoel Lubell, Paul Turner, Kristina Keitel

ABSTRACT

Introduction Early identification of children at risk of severe febrile illness can optimise referral, admission and treatment decisions, particularly in resource-limited settings. We aimed to identify prognostic clinical and laboratory factors that predict progression to severe disease in febrile children presenting from the community.

Methods We systematically reviewed publications retrieved from MEDLINE, Web of Science and Embase between 31 May 1999 and 30 April 2020, supplemented by hand search of reference lists and consultation with an expert Technical Advisory Panel. Studies evaluating prognostic factors or clinical prediction models in children presenting from the community with febrile illnesses were eligible. The primary outcome was any objective measure of disease severity ascertained within 30 days of enrolment. We calculated unadjusted likelihood ratios (LRs) for comparison of prognostic factors, and compared clinical prediction models using the area under the receiver operating characteristic curves (AUROCs). Risk of bias and applicability of studies were assessed using the Prediction Model Risk of Bias Assessment Tool and the Quality In Prognosis Studies tool.

Results Of 5949 articles identified, 18 studies evaluating 200 prognostic factors and 25 clinical prediction models in 24,530 children were included. Heterogeneity between studies precluded formal meta-analysis. Malnutrition (positive LR range 1.56–11.13), hypoxia (2.10–8.11), altered consciousness (1.24–14.02), and markers of acidosis (1.36–7.71) and poor peripheral perfusion (1.78–17.38) were the most common predictors of severe disease. Clinical prediction model performance varied—the best performing models being those evaluated in similar settings and using similar outcomes as the original derivation studies.

Conclusions Few studies address this important public health question. We identified prognostic factors from a wide range of geographic contexts that can help clinicians assess febrile children at risk of progressing to severe disease. Multicentre studies that include outpatients are required to explore generalisability and develop data-driven tools to support patient prioritisation and triage at the community level.

Key questions

What is already known?
- An increasing number of clinical decision-support algorithms and risk stratification tools integrate clinical and laboratory predictors to guide healthcare workers in their assessment of febrile children.
- Which prognostic factors—alone or as components of clinical prediction models—best identify children at risk of developing severe febrile illness is not clear.
- Previous systematic reviews have focused on diagnostic studies and used imperfect reference standards for severe disease.

What are the new findings?
- Malnutrition, hypoxia, altered consciousness, and bedside markers of acidosis and poor peripheral perfusion were the most commonly identified predictors of severe disease.
- Clinical prediction model performance varied—the best performing models being those evaluated in similar settings and using similar outcomes as the original derivation studies.
- The prognostic factors and clinical prediction models identified in this study reflect children with relatively advanced illnesses and hence the degree to which they can inform community triage and prioritisation strategies is unclear.

INTRODUCTION

Acute febrile illnesses are among the most common reasons that parents seek medical care for their children. While most episodes are mild, an important minority of children progress to severe disease. Early recognition of low-incidence serious disease is challenging, especially in many tropical settings where health workers receive limited training, patient volumes are high, diagnostic capacity is poor and different acute febrile syndromes are often clinically indistinguishable.
Clinical and laboratory prognostic factors that enable early and accurate identification of children at risk of developing severe disease could improve patient outcomes and reduce resource misallocation. An increasing number of clinical decision-support algorithms and risk stratification tools integrate clinical and laboratory predictors to guide referral, admission and treatment decisions. While no unified strategy exists to guide selection of candidate predictors, those already reported as prognostic should normally be considered.

Previous reviews have evaluated predictors of ‘serious bacterial infections’. However, these studies are diagnostic rather than prognostic. Furthermore, ‘serious bacterial infection’ is an imperfect measure of disease severity: microbiological tests for bacterial infections lack sensitivity, especially in settings with high antibiotic consumption; ‘serious bacterial infections’ are not always severe (eg, children with enteric fever are often successfully managed as outpatients) and severe febrile illnesses are frequently caused by non-bacterial pathogens, especially in low-middle-income countries (LMICs), in part secondary to the introduction of widespread vaccination against prevalent bacterial pathogens of childhood.

We performed a systematic review to identify which clinical and laboratory factors—alone or as part of clinical prediction models—predict progression to severe disease in febrile children presenting from the community to a community health worker, primary health center or hospital outpatient or emergency department. Our aim was to understand which prognostic factors might support health workers faced with this difficult and common clinical question and to inform variable selection for future prospective studies aiming to develop data-driven triage tools.

**METHODS**

**Protocol and registration**

The methods for this systematic review were specified in advance (PROSPERO protocol: CRD42019140542) and adhere to the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS), a modification of CHARMS for prognostic factor studies (CHARMS-PF), Quality In Prognosis Studies (QUIPS) and Prediction Model Risk of Bias Assessment Tool (PROBAST) guidelines. The report has been prepared in accordance with Preferred Reporting Items for Systematic Review and Meta-Analysis guidelines.

**Eligibility criteria**

All prognostic studies (prognostic factor and clinical prediction model) including ≥20 patients were eligible. Our target population was children aged >28 days and <19 years, presenting from the community with an acute febrile illness (documented abnormal temperature, fever or hypothermia) or history of fever or suspected sepsis. While sepsis is not always well defined in children, ‘suspected sepsis’ was included along with febrile children so as to include all children with suspected infection. Studies were excluded if disaggregated paediatric data were not presented or patients were recruited partway through receipt of inpatient treatment, as the aim of the review was to identify prognostic variables measured at presentation. Studies that only evaluated specific clinical syndromes (eg, neurological presentations, acute respiratory infections and so on) or particular pathogens (eg, *Plasmodium* spp, influenza and so on) were not included.

Studies measuring predictors at presentation to care were included. Studies where authors identified that a substantial proportion of participants were recruited following transfer from another health facility were excluded. Demographic, anthropometric, socioeconomic, clinical and historical variables were considered, as well as laboratory parameters measured at presentation to care. Studies only reporting variables that would not be available at the time of presentation to care (eg, blood culture results) were excluded.

The primary outcome was any objective measure of disease severity occurring within 30 days of measurement of the predictors or during hospitalisation. Studies assessing outcome at the same time point as baseline predictor measurements (diagnostic studies) were excluded.

**Search strategy and selection criteria**

We searched MEDLINE, Embase and Web of Science databases, without language restriction, for publications between 31 May 1999 and 30 April 2020 (initial search to 31 May 2019; updated search to 30 April 2020). We followed Cochrane Prognosis Methods Group recommendations to build our search strategy (online supplemental appendix S1), structured according to the ‘populations, interventions, comparators, outcomes, timing and setting’ (PICOTS) framework and adapted published search strings as appropriate. The search
strategy was peer-reviewed by an independent Technical Advisory Panel (online supplemental appendix S2).

Study selection
Title, abstract and full-text screening were performed independently by two reviewers (AC and RT). Agreement was checked after the first 20 and 250 articles. Discrepancies were resolved by discussion or independent assessment by a third reviewer (KK).

Eligible studies and relevant review articles were ‘snowballed’ (forward and reverse crosschecking of reference lists) to identify additional studies. The list of eligible studies was presented to the Technical Advisory Panel who were asked to identify obvious omissions and suggest key authors whose publication lists were subsequently reviewed for additional eligible studies (online supplemental appendix S2).

Data collection process
Data extraction sheets were developed based on the CHARMS and CHARMS-PF checklists (online supplemental appendix S3). Data were extracted independently by one reviewer (AC or RT) and checked by the other. Discrepancies were discussed and resolved between the two reviewers. Authors of studies not reporting likelihood ratios (LRs) (prognostic factors) or area under the receiver operating characteristic curves (AUROCs) (clinical prediction models), or the data to allow their calculation, were contacted. Seven authors responded to requests for clarifications and six provided additional data not available in the published manuscript. All predictors were harmonised using the Systematised Nomenclature of Medicine Clinical Terms (SNOMED-CT).

Data analysis: prognostic factors
Contingency tables were constructed and positive likelihood ratio (PLR) and negative likelihood ratio (NLR) calculated for each prognostic factor. In the case of an empty cell, 0.5 was added to each cell (Haldane-calculation, were harmonised using the Systematised Nomenclature of Medicine Clinical Terms (SNOMED-CT).

Data analysis: clinical prediction models
For clinical prediction models, AUROCs are presented on forest plots (Stata V.16.0). LRs were selected as the principal effect estimate as they allow estimation of post-test probabilities, are independent of prevalence, are intuitive for clinicians and are frequently used to compare performance of predictors in diagnostic and prognostic studies. Prognostic factors are presented in the main analysis if at least one study reported a PLR ≥5.0 (ie, a rule-in test), or a NLR ≤0.2 (ie, a rule-out test). To contextualise the results, we used the outcome prevalence of individual studies to calculate the pre-test probability, and display positive and negative post-test probabilities on dumbbell plots (R V.3.6.1).

Synthesis of results
Due to expected heterogeneity between studies (as a result of variations in case-mix and baseline risk), few common predictors for comparison and absence of well-defined subgroups, no formal meta-analysis nor comparison of variability and bias between studies was planned, as these comparisons are recognised as being prone to bias. Qualitative comparisons are described considering major differences between populations and study design. Prevalence of severe disease was used to group studies into low (<2.5%), moderate (2.5%–7.5%) and high (>7.5%) prevalence settings, as a proxy for the case-mix and level of care.

Quality assessment
Risk of bias and applicability of studies were assessed using the QUIPS tool for prognostic factor studies, and PROBAST for studies developing, validating or updating prediction models. Each study was independently assessed using QUIPS or PROBAST by two reviewers (AC and RT), as well as an independent senior reviewer (MC, AVDB or JV). All discrepancies were resolved by discussion. For prognostic factor studies (QUIPS), risk of bias was categorised as low, medium or high, while in clinical prediction model studies (PROBAST) risk was categorised as low, high or unclear. For all studies, applicability was assessed as being of high, low or unclear concern.

Role of the funding source
The funders had no role in study design, data collection, data analysis, data interpretation or writing of the report. The co-primary authors (AC and RT) had full access to the data and final responsibility for the decision to submit for publication.

RESULTS
The electronic search retrieved 5930 articles, and 19 additional articles were identified through snowballing and expert consultation (figure 1). Eighteen studies were included in the review: 16 studies evaluated 200 prognostic factors, from 75 SNOMED-CT categories, and eight evaluated 33 clinical prediction model/outcome pairs, using 25 distinct models.

In total 24 530 children were included, with overlap across eight studies. The majority (11/18) included only hospitalised patients. Two studies recruited children from primary care, and five included children admitted and those sent home directly from hospital outpatient or emergency departments. Seven studies included children aged 5 years and under, with the remainder including patients up to 19 years of age.
Figure 1  Selection of studies. Only one reason for exclusion per study is listed. CPM, clinical prediction model; PF, prognostic factor.

studies did not include a temperature measurement in their eligibility criteria and enrolled children on the basis of suspected infection or sepsis. Thirty-eight to forty-two Eight studies were conducted in sub-Saharan Africa, twenty-six to thirty-four four in North America, three in Europe, two in Asia, and one in Latin America. Six were multicentre studies. Most used ‘hard’ outcomes to define severe disease, such as mortality, organ dysfunction or need for organ support, while four used ‘softer’ outcomes, such as prolonged length of stay or persistence of symptoms. Characteristics of the 18 studies are summarised in Table 1.

Prognostic factors

Figures 2–4 present prognostic factors identified as having rule-in (PLR ≥5.0) or rule-out (NLR ≤0.2) value in at least one study. Prognostic factors that met neither of these pre-specified cut-offs are presented in online supplemental appendix S5. In settings with moderate prevalence of severe disease, both high lactate (PLR range 4.97–5.13) and hypoglycaemia (PLR range 12.63–13.36) were useful for ruling in severe disease, whereas a lactate ≤5 mM was more useful as a rule-out test (NLR 0.13) among a population in whom prevalence of severe disease was high (febrile children with signs of poor organ perfusion).

Hypoxia was most useful to rule-in severe disease in moderate prevalence settings (PLR range 8.11–9.49). Some studies found hypotension and bedside markers of poor peripheral perfusion to have useful rule-in value, but this was inconsistent (PLR range 1.89–9.57 and 1.78–17.38, respectively). Bradycardia was evaluated in a multicentre study conducted across three East African countries and found to have useful rule-in value (PLR range 5.95–14.59) for severe disease in those high prevalence settings. Impaired consciousness, assessed using bedside coma scales, was a useful predictor of severe disease, particularly in low and moderate prevalence settings (PLR range 3.38–14.02), with the post-test probability of poor outcome increasing with the degree of neurological impairment.

In sub-Saharan African settings, severe malnutrition (PLR range 1.56–11.25), HIV positive status (PLR range 2.32–12.48) and bedside correlates of metabolic derangement such as deep breathing and jaundice (PLR range 3.57–7.71) were useful rule-in predictors, across a range of prevalence settings.

Very few prognostic factors were satisfactorily able to rule-out progression to severe disease: presence of comorbidities (NLR range 0.12–1.04), sepsis at admission (NLR 0.19) and prostration (NLR range 0.18–1.23) were each identified in only one study.

Clinical prediction models

Figure 5 illustrates the discrimination (AUROC) of 25 clinical prediction models for 33 different outcomes assessed in eight studies: most (18/33) were external validations of existing models; thirteen were newly derived models and two were updates and external validations of an existing model. Components of the clinical prediction models are summarised in Table 2.

Three models, Lambaréné Organ Dysfunction Score (LODS), Paediatric Early Death Index for Africa (early death score) (PEDIA-e) and Signs of Inflammation in Children that Kill (SICK), showed good (AUROC ≥0.80) discrimination in a Ugandan setting where in-hospital mortality occurred at a prevalence of 4.7% (AUROC range 0.85–0.90). Two of these (LODS and PEDIA-e) were also assessed in a multicentre study in East Africa where discrimination was lower (AUROCs of 0.77 and 0.70). This study also derived two models, the FEAST-Paediatric Emergency Triage (FEAST-PET) and FEAST-Paediatric Emergency Triage and Laboratory (FEAST-PETaL) scores, which showed good discrimination (AUROCs of 0.86 and 0.82). Two other East African studies used combinations of simple clinico-demographic variables to derive a number of prediction models, four of which had AUROCs ≥0.80.

One North American study derived a model to predict hypotensive shock in unselected children presenting with suspected sepsis, which showed good discrimination in an external geographic validation (AUROC 0.87). The Yale Observation Score also showed high discrimination for mortality (AUROC 0.97) and mechanical ventilation (AUROC 0.89) in India, however, the small sample size (n=100) renders the results difficult to interpret. In general, models assessed against ‘softer’ outcomes (eg,
Table 1  Characteristics of included studies

<table>
<thead>
<tr>
<th>Study (year); setting, country</th>
<th>Cohort</th>
<th>Design</th>
<th>Quality assessment</th>
<th>CPM or PF study*</th>
<th>Sample size</th>
<th>Population</th>
<th>Outcome prevalence (n/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scott (2020)$^6$; Hospital OPD/ED</td>
<td>Retrospective cohort</td>
<td>High</td>
<td>Low</td>
<td>CPM</td>
<td>2464</td>
<td>Age 60 days to 18 years; Clinician-suspected sepsis</td>
<td>Hypotensive septic shock on arrival; transfer to another centre; leaving ED before formal evaluation; incorrect registration</td>
</tr>
<tr>
<td>Walia (2016)$^8$; Hospitalised§</td>
<td>Prospective cohort</td>
<td>High</td>
<td>High</td>
<td>CPM</td>
<td>100</td>
<td>Age 3–36 months; Axillary temperature &gt;36.9°C (early morning) or &gt;37.4°C</td>
<td>Non-infectious cause of fever; immunisation ≤2 days; immunodeficiency, autoimmune disorder</td>
</tr>
<tr>
<td>Aramburo (2018)$^9$; Hospitalised§</td>
<td>Randomised controlled trial</td>
<td>Moderate</td>
<td>High</td>
<td>PF</td>
<td>3008</td>
<td>Age 60 days to 12 years; history of fever or axillary temperature ≥37.5°C or &lt;36°C; severe febrile illness¶</td>
<td>Non-infectious cause of illness; SAM, gastroenteritis, burns, chronic kidney disease, pulmonary oedema, intoxication, surgical conditions, receipt of isotonic fluids during the same illness</td>
</tr>
<tr>
<td>George (2015)$^{10}$; Hospitalised§</td>
<td>Randomised controlled trial</td>
<td>High</td>
<td>High</td>
<td>CPM</td>
<td>3121</td>
<td>Age 60 days to 12 years; history of fever or axillary temperature ≥37.5°C or &lt;36°C; severe febrile illness¶</td>
<td>Non-infectious cause of illness; SAM, gastroenteritis, burns, chronic kidney disease, pulmonary oedema, intoxication, surgical conditions, receipt of isotonic fluids during the same illness</td>
</tr>
<tr>
<td>Scott (2012)$^7$; Hospital OPD/ED</td>
<td>Prospective cohort</td>
<td>High</td>
<td>High</td>
<td>PF</td>
<td>239</td>
<td>Age &lt;19 years; temperature &gt;38.5°C or &lt;36°C and heart rate &gt;2 SD above normal for age; underwent phlebotomy as part of usual care</td>
<td>Transfer from another health facility; known inborn errors of metabolism; receipt of &gt;15 min of intravenous therapy</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Study (year); setting, country</th>
<th>Population</th>
<th>Outcome prevalence (n/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scott (2014) Tertiary care hospital, USA</td>
<td>Age &lt;19 years; temperature &gt;38.5°C or &lt;36°C and heart rate &gt;2 SD above normal for age; undergoing phlebotomy as part of routine care</td>
<td>Transfer from another health facility; known inborn errors of metabolism; receipt of &gt;15 min of intravenous therapy</td>
</tr>
<tr>
<td>Nadjm (2013) Secondary care hospital, Tanzania</td>
<td>Age 2 months to 5 years; history of fever in last 48 hours or axillary temperature &gt;37.5°C</td>
<td>Chronic illness (excluding HIV and malnutrition); trauma; surgical conditions</td>
</tr>
<tr>
<td>Mtove (2011) Secondary care hospital, Tanzania</td>
<td>Age 2 months to 13 years; history of fever in last 48 hours or axillary temperature &gt;37.5°C</td>
<td>Chronic illness (excluding HIV and malnutrition); trauma; surgical conditions</td>
</tr>
<tr>
<td>Lowlaavar (2016) Secondary and tertiary care hospitals, Uganda</td>
<td>Age 6–60 months; admitted during study working hours or within 8 hours of study shift with a proven or suspected infection</td>
<td>Previous enrolment; residence outside study catchment area</td>
</tr>
<tr>
<td>Conroy (2015) Tertiary care hospital, Uganda</td>
<td>Age 2 months to 5 years; history of fever in last 48 hours or axillary temperature &gt;37.5°C</td>
<td>None reported</td>
</tr>
<tr>
<td>van Nassau (2018) Secondary care hospital, The Netherlands</td>
<td>Age &lt;18 years; suspected bacterial infection**</td>
<td>Surgical conditions</td>
</tr>
<tr>
<td>Study (year); setting, country</td>
<td>Cohort</td>
<td>Design</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td><strong>Scott (2017)</strong>&lt;sup&gt;10&lt;/sup&gt;; Tertiary care hospital, USA</td>
<td>Hospital OPD/ED</td>
<td>Retrospective cohort</td>
</tr>
<tr>
<td><strong>SEAIDCRN (2017)</strong>&lt;sup&gt;11&lt;/sup&gt;; Tertiary care hospitals, Indonesia, Thailand and Vietnam</td>
<td>Hospitalised§</td>
<td>Prospective cohort</td>
</tr>
<tr>
<td><strong>Costa de Santana (2017)</strong>&lt;sup&gt;12&lt;/sup&gt;; Tertiary care hospital, Brazil</td>
<td>Hospital OPD/ED</td>
<td>Retrospective cohort</td>
</tr>
<tr>
<td><strong>Kwizera (2019)</strong>&lt;sup&gt;13&lt;/sup&gt;; Secondary care hospital, Rwanda</td>
<td>Hospitalised§</td>
<td>Prospective cohort</td>
</tr>
</tbody>
</table>

**Outcomes including length of stay and persistence of symptoms**

| Freyne (2013)<sup>14</sup>; Secondary care hospital, Ireland | Hospitalised§ | Prospective cohort | Risk of bias Low | PF | 46 | Age 6–36 months; axillary temperature >38.1°C | Chronic illness; immunisation ≤2 days, antipyretic use ≤2 hours | Length of stay >96 hours | 26.1% (12/46) |

<sup>10</sup> Scott et al., <sup>11</sup> SEAIDCRN et al., <sup>12</sup> Costa de Santana et al., <sup>13</sup> Kwizera et al., <sup>14</sup> Freyne et al.
### Table 1 Continued

<table>
<thead>
<tr>
<th>Study (year); setting, country Cohort</th>
<th>Quality assessment</th>
<th>Population</th>
<th>Outcome prevalence (n/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>van Nassau (2018)</strong>§§; Secondary care hospital, The Netherlands</td>
<td>Hospitalised§</td>
<td>Retrospective cohort</td>
<td>High</td>
</tr>
<tr>
<td><strong>Elsbout (2015)</strong>§§; General Practice (out of hours), The Netherlands</td>
<td>Primary care</td>
<td>Prospective cohort</td>
<td>High</td>
</tr>
<tr>
<td><strong>Mwandama (2016)</strong>§§; Community Health Workers, Malawi</td>
<td>Primary care</td>
<td>Prospective cohort</td>
<td>High</td>
</tr>
</tbody>
</table>

Studies are grouped according to the type of outcome they used: ‘hard’ (death, organ dysfunction, organ support, PICU admission) or ‘soft’ (length of stay, persistence of symptoms).

*Studies evaluating both PFs and CPMs were categorised on the basis of their primary analysis to facilitate review using the appropriate quality assessment tool.

† Hypotensive systolic blood pressure on arrival with receipt of a fluid bolus or vasoactive agent within 30 min.

‡ Hypotension plus receipt of >30 mL/kg isotonic crystalloids or vasoactive medication.

§ Only children the treating physician decided to admit were eligible but recruitment occurred at the time of admission to the health facility.

¶ Respiratory distress (increased work of breathing or deep breathing and/or impaired consciousness (coma or prostration) AND evidence of poor peripheral perfusion (capillary refill time >2 s or lower limb temperature gradient or weak radial pulse or severe tachycardia).

** Initiation of antibiotics within 24 hours of arrival in the emergency department.

†† Decreased mental status or perfusion in the setting of suspected infection.

‡‡ Rectal temperature >38.5°C or <35°C (or equivalent) AND heart rate >2 SD above normal for age (unless hypothermic) AND respiratory rate >2 SD above normal for age AND altered mental status OR systolic blood pressure <2 SD below normal for age OR pulse pressure <20 mm Hg OR capillary refill time >2 s OR SpO₂ <95% OR leucocyte count >12×10³ cells/µL or <5×10³ cells/µL.

§§ Convulsions, repeated vomiting, lethargy, severe anaemia or loss of consciousness.

CPM, clinical prediction model; ED, emergency department; n, number of outcomes; n, number of cases; OPD, outpatient department; PF, prognostic factor; PICU, paediatric intensive care unit; SAM, severe acute malnutrition; SEAIDCRN, Southeast Asia Infectious Disease Clinical Research Network; SIRS, systemic inflammatory response syndrome;
Figure 2  Prognostic factors identified as having rule-in (PLR ≥5.0) or rule-out (NLR ≤0.2) value for severe disease in at least one study—laboratory tests. mM, millimolar; NLR, negative likelihood ratio; PICU, paediatric intensive care unit; PLR, positive likelihood ratio.

Figure 3  Prognostic factors identified as having rule-in (PLR ≥5.0) or rule-out (NLR ≤0.2) value for severe disease in at least one study—cardiovascular, respiratory or neurological signs. In the study by Costa et al., ‘sepsis’ was defined according to the systemic inflammatory response syndrome (SIRS), requiring measurement of heart rate, respiratory rate, temperature and leucocyte count. In the study by Kwizera et al., ‘sepsis’ was defined according to the qSOFA Score in children aged ≥15 years, and using a combination of temperature, mental status, respiratory distress, prostration and seizures in children aged <15 years. AVPU, alert, voice, pain or unresponsive; BCS, Blantyre Coma Score; bpm, beats per minute; CRT, capillary refill time; GCS, Glasgow Coma Score; HR, heart rate; LLTG, lower limb temperature gradient; NLR, negative likelihood ratio; PICU, paediatric intensive care unit; PLR, positive likelihood ratio; qSOFA, quick Sequential Organ Failure Assessment.
Figure 4  Prognostic factors identified as having rule-in (PLR ≥5.0) or rule-out (NLR ≤0.2) value for severe disease in at least one study—historical, anthropometric and metabolic variables. Children with visible wasting or nutritional oedema were also classified as having severe malnutrition. In the study by Elshout et al., ‘comorbidity’ was defined as being under routine care of a paediatrician or ENT specialist. ENT, ear, nose and throat; MUAC, mid-upper arm circumference; NLR, negative likelihood ratio; PICU, paediatric intensive care unit; PLR, positive likelihood ratio; WAZ, weight-for-age z-score.

Figure 5  Discrimination of clinical prediction models to identify children at risk of severe disease. Individual studies evaluated different clinical prediction models using datasets with varying numbers of children with severe disease, depending on the data available. The outcome prevalence reflects the proportion of children with severe disease in the dataset used to evaluate that particular prediction model/outcome pair. This may be different from the overall prevalence of children with severe disease in the study, which is listed in table 1 and used to classify studies into low, moderate or high prevalence settings. No CIs were provided for the AUROC estimates in the study by Walia et al. AQUAMAT, African Quinine Artesunate Malaria Trial; AUROC, area under the receiver operating characteristic curve; FEAST-PET, FEAST-Paediatric Emergency Triage; FEAST-PETaL, FEAST-Paediatric Emergency Triage and Laboratory; LODS, Lambaréné Organ Dysfunction Score; PEDIA, Paediatric Early Death Index for Africa; PEWS, Paediatric Early Warning Score; PICU, paediatric intensive care unit; PRISM III, Paediatric Risk of Mortality; qPELOD-2, quick Paediatric Logistic Organ Dysfunction; qSOFA, quick Sequential Organ Failure Assessment; SICK, Signs of Inflammation in Children that Kill; SIRS, Systemic Inflammatory Response Syndrome; YOS, Yale Observation Score.
<table>
<thead>
<tr>
<th>Clinical prediction model</th>
<th>Variables used in the clinical prediction model in the included studies</th>
<th>Included study evaluating the model</th>
<th>Original study developing the model</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQUAMAT</td>
<td>Base deficit, impaired consciousness, convulsions, elevated blood urea, underlying chronic illness</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;</td>
<td>von Seidlein&lt;sup&gt;60&lt;/sup&gt;</td>
</tr>
<tr>
<td>ELSHOUT model</td>
<td>Sore throat, palpable lymph nodes, duration of fever before consultation, C-reactive protein</td>
<td>Elshout&lt;sup&gt;29&lt;/sup&gt;</td>
<td>Elshout&lt;sup&gt;29&lt;/sup&gt;</td>
</tr>
<tr>
<td>FEAST-PET</td>
<td>Axillary temperature, heart rate, capillary refill time, conscious level, respiratory distress, lung crepitations, severe pallor, weak pulse</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;</td>
</tr>
<tr>
<td>FEAST-PETaL</td>
<td>FEAST-PET with the addition of lactate, pH, blood urea nitrogen</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;</td>
</tr>
<tr>
<td>KWIZERA model 1</td>
<td>Age, respiratory rate, heart rate, temperature, capillary refill time, altered mental state</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
</tr>
<tr>
<td>KWIZERA model 2</td>
<td>Age, respiratory rate, heart rate, capillary refill time, altered mental state</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
</tr>
<tr>
<td>KWIZERA model 3</td>
<td>Age, respiratory rate, temperature, capillary refill time, altered mental state</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
</tr>
<tr>
<td>KWIZERA model 4</td>
<td>Age, respiratory rate, capillary refill time, altered mental state</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
</tr>
<tr>
<td>KWIZERA model 5</td>
<td>Age, respiratory rate, altered mental state</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
<td>Kwizera&lt;sup&gt;41&lt;/sup&gt;</td>
</tr>
<tr>
<td>LODS</td>
<td>Deep breathing, coma, and prostration</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;; Conroy&lt;sup&gt;27&lt;/sup&gt;</td>
<td>Helbok&lt;sup&gt;53&lt;/sup&gt;</td>
</tr>
<tr>
<td>LOWLAAVAR model 1</td>
<td>Conscious level, HIV, weight-for-age z-score</td>
<td>Lowlaavar&lt;sup&gt;42&lt;/sup&gt;</td>
<td>Lowlaavar&lt;sup&gt;42&lt;/sup&gt;</td>
</tr>
<tr>
<td>LOWLAAVAR model 2</td>
<td>Conscious level, HIV, mid-upper arm circumference</td>
<td>Lowlaavar&lt;sup&gt;42&lt;/sup&gt;</td>
<td>Lowlaavar&lt;sup&gt;42&lt;/sup&gt;</td>
</tr>
<tr>
<td>LOWLAAVAR model 3</td>
<td>Conscious level, mid-upper arm circumference</td>
<td>Lowlaavar&lt;sup&gt;42&lt;/sup&gt;</td>
<td>Lowlaavar&lt;sup&gt;42&lt;/sup&gt;</td>
</tr>
<tr>
<td>PEDIA-i</td>
<td>Anaemia, jaundice, indrawing, deep breathing, conscious level, prostration, convulsions/seizures, temperature</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;</td>
<td>Berkley&lt;sup&gt;51&lt;/sup&gt;</td>
</tr>
<tr>
<td>PEDIA-e</td>
<td>Jaundice, indrawing, conscious level, prostration, convulsions/seizures, wasting, kwashiorkor&lt;sup&gt;*&lt;/sup&gt;</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;; Conroy&lt;sup&gt;27&lt;/sup&gt;</td>
<td>Berkley&lt;sup&gt;51&lt;/sup&gt;</td>
</tr>
<tr>
<td>PEDIA-I</td>
<td>History &gt;7 days, conscious level, prostration, convulsions/seizures, temperature, wasting, kwashiorkor&lt;sup&gt;*&lt;/sup&gt;</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;</td>
<td>Berkley&lt;sup&gt;51&lt;/sup&gt;</td>
</tr>
<tr>
<td>PEWS†</td>
<td>Heart rate, capillary refill time, respiratory rate, oxygen saturation, systolic blood pressure</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;</td>
<td>Parshuram&lt;sup&gt;31&lt;/sup&gt;</td>
</tr>
<tr>
<td>PRISM III†</td>
<td>Heart rate, temperature, conscious level, systolic blood pressure, glucose, potassium, PCO₂, pH, acidosis, pupillary reflexes</td>
<td>George&lt;sup&gt;31&lt;/sup&gt;</td>
<td>Pollack&lt;sup&gt;62&lt;/sup&gt;</td>
</tr>
<tr>
<td>qPELOD-2</td>
<td>Systolic or mean arterial pressure, heart rate, altered mentation</td>
<td>van Nassau&lt;sup&gt;38&lt;/sup&gt;</td>
<td>Leclerc&lt;sup&gt;53&lt;/sup&gt;</td>
</tr>
<tr>
<td>qSOFA</td>
<td>Respiratory rate, altered mentation, systolic blood pressure</td>
<td>van Nassau&lt;sup&gt;38&lt;/sup&gt;</td>
<td>Seymour&lt;sup&gt;54&lt;/sup&gt;</td>
</tr>
<tr>
<td>qSOFA-L</td>
<td>qSOFA with the addition of lactate</td>
<td>van Nassau&lt;sup&gt;38&lt;/sup&gt;</td>
<td>van Nassau&lt;sup&gt;38&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Continued
of bias (figure 6). Most often this was due to inadequate reporting of model performance (studies reporting discrimination but not calibration), circularity between predictors and outcomes or having fewer than 100 participants with severe outcomes for model validation. It is noteworthy that one study which externally validated three models included 99 children who died. Another study which derived and/or validated nine models undertook an additional external validation in a population of acutely unwell but non-febrile children (and thus not eligible for consideration in this review), which included more than 100 children who died.

In all but one study there was high concern regarding applicability to the review question. This was largely due to the majority of studies including only children requiring hospitalisation, with recruitment occurring after the decision to admit had been made by the treating physician. Full details on risk of bias and applicability assessments are provided in online supplemental appendix S6.

**DISCUSSION**

This systematic review of prognostic factors and clinical prediction models assessing severity of disease in febrile children highlights that few well-conducted studies address this important public health question, particularly in unselected children presenting from the community. One of its main strengths is the inclusion of studies from a wide geographic context, aiding understanding of how predictive performance can vary across settings. By focusing on prognosis, we identified features that predict the likelihood that a child’s illness might progress, rather than features associated with illness severity at the moment of assessment.

Most prognostic factors identified as valuable for predicting severe childhood febrile illness (PLR ≥ 5.0) overlapped with individual components of the most...
promising clinical prediction models (AUROC ≥0.80): nutritional and HIV status, hypoxia, altered consciousness, and markers of acidosis (raised venous lactate or deep breathing) and poor peripheral perfusion (weak pulse, limb-core temperature gradient or prolonged capillary refill time).4-2 Hypoglycaemia was a useful prognostic factor identified in our review, but omitted in most clinical prediction models. Many of these features, however, indicate a child that is already very unwell, reflecting the fact that most studies included only hospitalised children and focused on predicting mortality. Few prognostic factors adequately ruled out (NLR ≤0.2) the possibility of progression to severe disease, a finding consistent with a previous systematic review evaluating the diagnostic utility of clinical features for serious bacterial infections.10

The major limitation of our work arises from the heterogeneity of studies, which precludes comparison of effect estimates. Second, it is difficult to determine if studies included children presenting to first-line health workers. We did not exclude studies solely based on the designated ‘level’ of a health facility: concerned parents in all settings use primary, secondary and tertiary care facilities as their first point-of-access. Third, most studies included only hospitalised children. This is a major barrier to understanding the potential for prognostic factors and prediction models to guide referral or admission decisions. Follow-up of children assessed as ‘low-risk’ (ie, those managed in the community) must be a priority for future studies seeking to determine the validity of prognostic factors and prediction models in outpatient settings.43 Fourth, in line with other reviews we found most studies to be of low quality.44 Recent guidance may help address this.17 Finally, we framed the review around ‘febrile illness’, rather than, for example, ‘clinically-suspected infection’. Our rationale was to ensure the findings were as relevant as possible for lesser-trained community health workers in resource-constrained settings, for whom a presumptive diagnosis of suspected infection can be challenging. Febrile illness is an accepted ‘pragmatic point-of-entry’ in these settings,45 however, we acknowledge that some children (particularly younger infants) may not mount a fever in response to serious infection. Therefore, despite our deliberately broad definition of febrile illness (documented abnormal temperature and history of fever), and the inclusion of studies of children with ‘suspected sepsis’, relevant studies may have been missed. Of note, in view of a suggestion arising during the peer-review process we also performed a second MEDLINE search, using alternate search strings, which did not yield any additional eligible articles (online supplemental appendix S7).

Thirty out of 200 (15%) prognostic factors met our pre-specified threshold for clinical relevance (PLR ≥5.0 or NLR ≤0.2). This may reflect the difficulty of identifying parsimonious predictors for all febrile children. While common pathophysiological pathways for severe disease have been identified across a spectrum of microbial aetiologies,46,47 certain predictors may perform better for specific syndromes or pathogens, compared with all-cause febrile illness. Five studies in our review reported a high proportion of children as being either slide-positive or rapid diagnostic test-positive for malaria. Notwithstanding the issues of co-infection and/or concomitant incidental parasitaemia in settings of high malaria endemicity, it is possible that the findings of these studies are more pertinent to children with malaria. However, four of these studies compared the prognostic performances of hyperlactaemia, hypoglycaemia and the prediction models SICK, LODS and PEDIA, and found them to be broadly consistent between children with malaria, non-malarial fever and invasive bacterial disease.26,27,32,34 Furthermore, as can be seen in figures 2–4, a number of predictors identified in malaria endemic regions also demonstrated prognostic utility in contexts where malaria is not endemic (eg, venous lactate, impaired peripheral perfusion, hypotension and altered consciousness). This, in conjunction with the subgroup analyses performed in the original studies, gives us a degree of confidence that the prognostic factors that we have identified are generalisable across different infecting pathogens. Nonetheless, future reviews using search strategies developed to retrieve syndrome-specific or pathogen-specific studies should explore this.

Another potential explanation for the relatively few valuable prognostic factors identified is work-up bias. In most studies, predictors were available to the treating clinicians: abnormal values are likely to have been acted on and predictive performance underestimated. For most predictors, particularly clinical signs, this is unavoidable as blindness is often neither possible nor ethical. When feasible, randomisation is required to definitively assess their potential impact.48 This is particularly important for new tests proposed in resource-limited settings. For example, both lactate and hypoxia were identified as potentially of value in this review but introducing tests for these parameters at all first-line health facilities across the tropics would incur substantial cost, and as their predictive value may vary in different settings, could result in unnecessary or missed referrals. Randomisation can help determine whether new tests such as these add value to simple clinical assessment.49

Clinical prediction models performed better when derived and validated in similar populations27: in East Africa LODS and PEDIA-ε (both derived in sub-Saharan Africa)50,51 were superior to SICK (originally derived in India).52 Model performance also improved when predicting the same outcome as the derivation study: quick Sequential Organ Failure Assessment and quick Paediatric Logistic Organ Dysfunction, derived to predict mortality, performed poorly when predicting prolonged length of stay.58,53,54 These findings highlight the importance of deriving prediction models using populations and outcomes appropriate to the clinical question. While mortality is a ‘hard’ outcome, it seldom occurs in primary care. Furthermore, its reflection of disease
severity is influenced (mediated) by the level of care. It is striking that in Tanzania a raised lactate conveyed a post-test probability of in-hospital mortality comparable to that of ‘organ dysfunction within 24 hours of arrival’ in a similar prevalence setting in the USA. Rather than relying on models derived in secondary care to generalise to outpatient settings across different epidemiological landscapes, alternative ways to quantify disease severity, which consider local context yet avoid circularity between predictor variables and outcome definitions, will be important to facilitate comparisons across settings and explore generalisability of risk prediction tools. Finally, the fact that most studies summarised model performance using only the AUROC means that is difficult to appreciate what the impact might be on clinical decision making.

In LMIC primary care contexts, many variables are not feasible to collect, and as noted above, some may incur substantial cost. Interestingly, HIV and nutritional status were both identified in our review and represent the only prognostic factors meeting our threshold for clinical relevance that may not necessarily reflect a child that is overtly very unwell. While biological plausibility for the prognostic utility of these two variables is high, it should be noted that the study which identified them was small and correspondingly the CI for the PLR is wide. The WHO’s Integrated Management of Childhood Illnesses ‘Danger Signs’ are recommended to guide referrals from community healthcare providers in resource-constrained settings. Of these, only altered consciousness was widely represented among included studies, and most found it to be a good predictor of severe disease. History of convulsions was examined in two studies while other ‘Danger Signs’ were not evaluated.

CONCLUSION
Our findings emphasise the limitations of individual prognostic factors. Performance varies widely across settings and clinicians must be cognisant not to over interpret individual predictors. While prediction models can support clinical decision making, they must be derived and validated using appropriate methodology, and populations and outcomes relevant to the clinical problem. For the identification of children at risk of severe febrile illness, this will require multiple, large, collaborative, research initiatives across different settings, which collect harmonised data on predictors and outcomes, and include unselected children presenting from the community.

Author affiliations
1Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia
2Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK
3Université Libre de Bruxelles, Brussels, Belgium
4University of Basel, Basel, Switzerland
5Swiss Tropical and Public Health Institute, Basel, Switzerland
6Department of Women and Children’s Health, King’s College London, London, UK
7Academic Centre of General Practice, University of Leuven, Leuven, Flanders, Belgium
8Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
9Ifakara Health Institute, Dar-es-Salaam, Tanzania
10Department of Pediatrics and Child Health, Muhimbili University Health and Allied Sciences, Dar-es-Salaam, Tanzania
11Mahidol-Oxford Tropical Medicine Research Unit, Bangkok, Thailand
12Division of Emergency Medicine, Department of Pediatrics, University Children’s Hospital, Inselhospital, University of Bern, Bern, Switzerland

Acknowledgements
We acknowledge the support of Nia Roberts and Jolanda Elmers in helping to translate MEDLINE search terms and access certain articles. We are grateful to Hanne Boon for assisting with the R code to create the dumbbell plots.

Contributors
AC conceived the study; AC, RT, YL, PT and KK defined the review strategy; AC and RT conducted the search, screened retrieved articles and extracted the data; AC, RT, MC, AVDB and JV assessed quality of included articles; AC and RT analysed the data and drafted the report; CK provided statistical oversight; PT and KK commented on the drafted report; AC, RT, MC, AVDB, JV, CK, NS, YL, PT and KK commented on and approved the final manuscript.

Funding
AC is supported by a Wellcome Trust Doctoral Training Fellowship. RT is supported by the Botnar Foundation. The Cambodia Oxford Medical Research Unit and Mahidol-Oxford Tropical Medicine Research Unit are part of the Wellcome Trust Thailand Africa and Asia Programme, which receives core-funding from the UK Wellcome Trust (106 698/14/2). The project also received seed funding from the Tropical Health Education Trust in the UK.

Competing interests
None declared.

Patient consent for publication
Not required.

Data availability statement
Data are available upon request. All data relevant to the study are included in the article or uploaded as supplementary information. The protocol for the study is available from: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=140542.

Supplemental material
This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and omissions arising from translation and adaptation or otherwise.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/

Author note
Technical Advisory Panel members: Jalemba Aluvaala, Quique Bassat, David Bell, John Crump, W. Conrad Liles, Rianne Oostenbrink and Shunmay Yeung. Affiliations are listed in online supplemental appendix S2.

ORCID iDs
Arjun Chandna http://orcid.org/0000-0003-1313-7922
Paul Turner http://orcid.org/0000-0002-1013-7815

REFERENCES


16 Moons KGM, Wolf RF, Riley RD. Searching for clinical prediction rules in MEDLINE to enhance systematic reviews. *PloS One* 2012;7:e39844.


43 Jaeschke R, Guyatt GH, Sackett DL. Users’ guides to the medical literature. III. How to use an article about a diagnostic test. B. what are the results and will they help me in caring for my patients? the evidence-based medicine Working group. *JAMA* 1994;271:703–7.


