BMC Medicine

How well do clinical prediction rules perform in identifying serious infections in acutely ill children across an international network of ambulatory care datasets?

Jan Y Verbakel[1*], Ann Van den Bruel[2], Matthew Thompson[2], Richard Stevens[2], Bert Aertgeerts[1], Rianne Oostenbrink[3], Henriette A Moll[3], Marjolein Y Berger[4], Monica Lakhanpaul[5], David Mant[2] and Frank Buntinx[1,6], for the European Research Network on Recognising Serious Infection (ERNIE)

Abstract

Background: Diagnosing serious infections in children is challenging, because of the low incidence of such infections and their non-specific presentation early in the course of illness. Prediction rules are promoted as a means to improve recognition of serious infections. A recent systematic review identified seven clinical prediction rules, of which only one had been prospectively validated, calling into question their appropriateness for clinical practice. We aimed to examine the diagnostic accuracy of these rules in multiple ambulatory care populations in Europe.

Methods: Four clinical prediction rules and two national guidelines, based on signs and symptoms, were validated retrospectively in seven individual patient datasets from primary care and emergency departments, comprising 11,023 children from the UK, the Netherlands, and Belgium. The accuracy of each rule was tested, with pre-test and post-test probabilities displayed using dumbbell plots, with serious infection settings stratified as low prevalence (LP; <5%), intermediate prevalence (IP; 5 to 20%), and high prevalence (HP; >20%) . In LP and IP settings, sensitivity should be >90% for effective ruling out infection.

Results: In LP settings, a five-stage decision tree and a pneumonia rule had sensitivities of >90% (at a negative likelihood ratio (NLR) of < 0.2) for ruling out serious infections, whereas the sensitivities of a meningitis rule and the Yale Observation Scale (YOS) varied widely, between 33 and 100%. In IP settings, the five-stage decision tree, the pneumonia rule, and YOS had sensitivities between 22 and 88%, with NLR ranging from 0.3 to 0.8. In an HP setting, the five-stage decision tree provided a sensitivity of 23%. In LP or IP settings, the sensitivities of the National Institute for Clinical Excellence guideline for feverish illness and the Dutch College of General Practitioners alarm symptoms ranged from 81 to 100%.

Conclusions: None of the clinical prediction rules examined in this study provided perfect diagnostic accuracy. In LP or IP settings, prediction rules and evidence-based guidelines had high sensitivity, providing promising rule-out value for serious infections in these datasets, although all had a percentage of residual uncertainty. Additional clinical assessment or testing such as point-of-care laboratory tests may be needed to increase clinical certainty. None of the prediction rules identified seemed to be valuable for HP settings such as emergency departments.

Keywords: clinical prediction rules, serious infection in children, external validation, NICE guidelines feverish illness, Yale Observation Scale, diagnostic accuracy

* Correspondence: jan.verbakel@med.kuleuven.be
[1]Department of General Practice, KU Leuven, Kapucijnenvoer 33 blok J, 3000 Leuven, Belgium
Full list of author information is available at the end of the article

Background

Acute infection is the most common presentation in children attending settings of ambulatory care (AC) [1,2]. Although most infections are self-limiting, they remain an important cause of morbidity and mortality in children in economically developed countries [3-5]. In the UK, infections account for 20% of childhood deaths, especially in children under 5 years of age [6]. Serious infections in children are usually defined as sepsis (including bacteremia), meningitis, pneumonia, osteomyelitis, cellulitis, and complicated urinary-tract infection (UTI; positive urine culture combined with systemic features such as fever) [3]. As a result of immunization against *Haemophilus influenzae* and *Streptococcus pneumoniae*, the incidence of these diseases has decreased steadily over recent decades, and they are now estimated to account for less than 1% of all acute childhood infections in primary care (PC) [2,7].

The combination of low incidence, non-specific initial clinical presentation, and potential for rapid deterioration makes the assessment of acutely ill children difficult [8,9]. Clinical prediction rules (CPRs) and guidelines may assist in the early recognition of serious infections [3]. In a previous systematic review, we identified all available CPRs (seven in total), based on signs and symptoms, for identifying any serious infection (two rules), pneumonia (two), meningitis (two), and dehydration from gastroenteritis (one rule) in AC settings [3]. Four of these seven CPRs were derived for use in emergency-care settings and their applicability in PC and AC settings has not been confirmed.

Only one rule, the Yale Observation Scale (YOS) [10] has been prospectively assessed in four studies [11-14], of which only two assessed the YOS in the intended age group of 3 to 36 months [12,14]. We also identified two national guidelines for the assessment of feverish children (Guideline on Feverish Illness in Children by the National Institute for Health and Clinical Excellence (NICE) [15] and the guidelines from the Dutch College of General Practitioners (NHG) [16]). A focused literature search identified an additional CPR published after this review: an emergency-department (ED) rule [17] to diagnose pneumonia, UTI, or bacteremia (see Additional file 1).

Although some of these guidelines (NICE guidelines, NHG alarm symptoms) are often used in clinical practice, very little external validation to support their use in practice has been performed in new and independent populations [18]. This raises questions about the robustness of the rules and their generalizability.

The aim of this study was to examine the diagnostic accuracy both of the CPRs identified by the systematic review and of the evidence-based guidelines, using retrospective external validations on individual patient datasets from ambulatory pediatric settings including PC and ED settings from three European countries.

Methods

Identification of datasets

We included datasets from studies identified in the systematic review [3], which had been published within the past 10 years, and from expert contacts. The criteria used to select datasets (Table 1), were design (cohort studies that enrolled children consecutively), sample size (> 500 children), participants (children aged 0 to 18 years or subgroups of these), setting (AC defined as general or family practice, pediatric outpatient clinics, pediatric assessment units, or EDs in developed countries), outcome (serious infection), and data availability (agreement to share data) (Figure 1).

Ethics approval

This research conformed to the Helsinki Declaration and to local legislation. The study authors, agreeing to share data, obtained ethics approval from their regional research ethics committees before the study for the initial data collection of the included datasets.

Processing of included datasets

Direct access to the raw data of each dataset was granted and key characteristics of each of the datasets were extracted (Table 2). The variables used in each dataset were translated to English if necessary, and the translation, coding, and definition of variables were clarified with the authors of the relevant study.

We used the following criteria to determine which dataset could be used to validate each CPR and guideline, and which diagnoses should be included in the composite outcome of serious infection.

• Datasets used to derive a CPR were not used to validate the same rule.

• When variables were not entirely identical with the variables of the original CPR, we identified proxies where possible. For example, the variable 'dyspnea' of the five-stage decision tree (FSDT) and the pneumonia rule was not recorded in three datasets; we therefore used either 'respiratory distress' or 'chest flaring' as a proxy (for a full list of all approximations, see Additional file 2 and Additional file 3).

• Based on the number of required variables, whenever one-third or more (fever guidelines), one or more (pneumonia rule, meningitis rule) or two or more (YOS, FSDT) of the required variables were not recorded, that dataset was not used for validation of that specific rule. We performed sensitivity analyses as described below.

• Missing data on variables used in the validation were not imputed because the necessary missing-at-random assumption was likely to be incorrect because some of

Table 1 Criteria for inclusion and exclusion of datasets in validation analysis.

Characteristic	Inclusion	Exclusion
Publication date	Studies published in the past 10 years	Studies published before 2003
Design	Studies that had recorded clinical features; prospective or retrospective cohort study design	Unclear methods
Sample Size	> 500 children	< 500 children
Participants	Age between 1 month and 18 years of age; studies including children spanning this age range were included if they reported age (or age could be calculated)	Children with congenital or obtained immunodeficiency. Age outside the required range
Setting	Ambulatory care (defined as general or family practice, pediatric outpatient clinics, pediatric assessment units, or emergency departments). Developed countries, defined using the United Nations list, which included Europe, Canada, USA, Australia, New Zealand and Japan	Studies conducted in developing countries
Outcome	Serious infection, defined as sepsis (including bacteremia), meningitis, pneumonia, osteomyelitis, cellulitis, and complicated urinary-tract infection (positive urine culture and systemic effects such as fever)	Diagnosis other than serious infection
Data availability	Agreement to share data	

the datasets consisted of routinely collected data from medical records.

• Apart from the approximations used (see Additional files 2, Additional file 3), no alterations of the original data were performed. We report the number of observations available for analysis of each prediction rule after applying these assumptions.

• In contrast to the other dichotomous rules, the YOS generates a sum score. We defined an abnormal result using two pre-selected cut-offs (of 8 or 10).

• Serious infection was defined as sepsis (including bacteremia), meningitis, pneumonia, osteomyelitis, cellulitis, or complicated UTI [3]. These diagnoses were available for all datasets, and assessment of the diagnoses to ensure comparability of outcomes was discussed with the authors of each study.

The settings in the included datasets were stratified as having low prevalence (LP; 0 to 5%), intermediate prevalence (IP; 5 to 20%) or high prevalence (HP; >20%) of the serious infection(s) of interest (including all serious infections, pneumonia, meningitis) with the clinical assumption that diagnostic goals are different in each setting. In LP settings, CPRs should have high sensitivity in order to correctly rule out (at a negative likelihood ratio (NLR) of up to 0.2) the target disorder(s) at a reasonable cost in terms of referral or admission rates [19,20].

The accuracy of the CPRs was assessed retrospectively in each of the available prospectively collected datasets by calculating sensitivity, specificity, predictive value, and likelihood ratio (LR). We used dumbbell plots to display the change from pre-test to post-test probabilities [3].

To avoid the risk of influencing diagnostic accuracy by either an arbitrarily chosen number of required variables, or the age range available in each dataset compared with the intended age range of the rule, we performed the

following sensitivity analyses after obtaining initial results with the different CPRs. Firstly, when a CPR was specifically designed for a certain age group (, for example, the YOS for children aged 3 to 36 months and the NICE guidelines for children up to 5 years of age), we compared the 95% confidence intervals (CIs) of the diagnostic characteristics (sensitivity, specificity, LRs and area under the curve (AUC)) [21] in the target age group with the entire age range of the dataset at hand. Second, when one or more variables of the original prediction rule were missing, we examined those same diagnostic characteristics in the datasets with no missing variables, to avoid biasing results on the number of missing variables. Whenever more than one (for the CPRs) or more than two (for the fever guidelines) original variables were missing, we did not perform sensitivity analysis, based on the rationale that missing two (or more) of a maximum of six variables (for the CPRs) or three (or more) of a maximum of eight original variables (for the fever guidelines) did not seem clinically sensible. This was discussed and confirmed by all study authors, contributing data to the current study.

Meta-analysis of the pooled results of the multiple external validations was not possible because substantial clinical heterogeneity was found in these datasets, including differences in setting, inclusion criteria, immunization schedules, and definition of serious infection. Additionally, the small number of included studies would have led to a high level of uncertainty in the estimates of the variances of the random effects for both the bivariate and hierarchical summary receiver operating characteristic models, if heterogeneity were to be explored statistically. Inclusion or exclusion of a single study would affect the convergence of the model greatly [21]. The individual patient data were analyzed in every dataset separately. The translation, re-coding, and data checking were performed by one author (JV), and the

Figure 1 Flowchart of dataset inclusion.
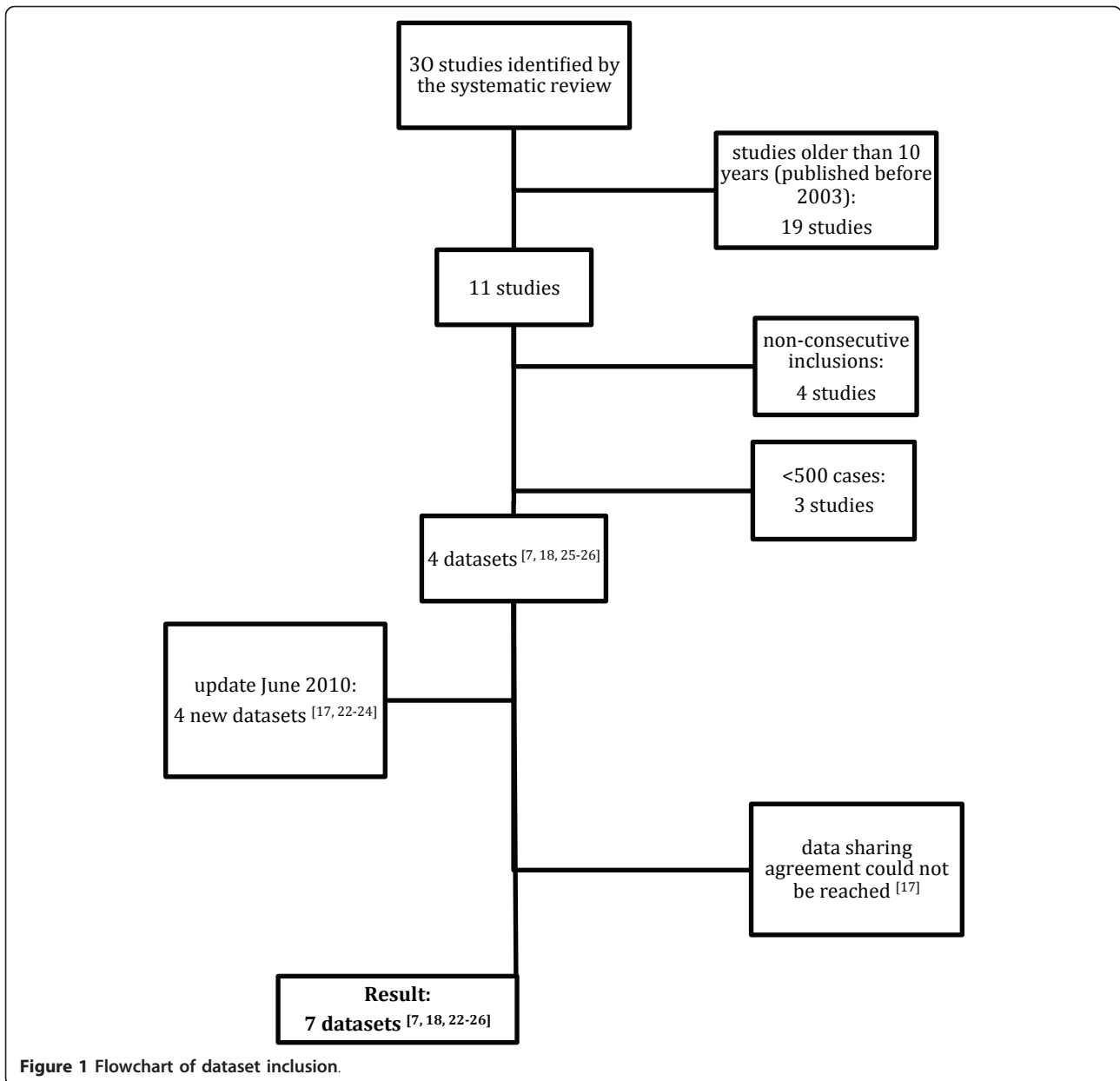
results of each step were discussed with all of the other authors. All analyses were performed with Stata software (version 11.2; Stata Corp., College Station, TX, USA).

Results
Included datasets
We obtained seven datasets providing data on 11,023 children: two LP datasets from general practice [7,22], two IP datasets from EDs [23,24] and three HP datasets from EDs [25,26] or pediatric assessment unit s[18] in the UK (n = 2), the Netherlands (n = 4) and Belgium (n = 1) (Figure 1, Table 2). Children were included based on presence of fever [22,24,25], acute illness [7,23], or acute infection

[18], or on referral for meningeal signs [26]. Children with various co-morbidities were excluded in six studies, and one study excluded children who required immediate resuscitation. The outcome in all studies included sepsis, meningitis, pneumonia, and complicated UTI as part of the outcome variables. Osteomyelitis and cellulitis were explicitly mentioned in five and three datasets, respectively. The mean age ranged from 0.94 to 5.0 years, and prevalence of serious infection ranged from 0.8 to 43.8%.

Clinical predictors included in the datasets
Most datasets included basic demographic characteristics such as age, duration, and severity of illness, as well

Table 2 Characteristics of datasets used for external validation of prediction rules.

Dataset	Setting	Country	N	Age, years; mean (range)	Prevalence ofserious infection % (95% CI)	Inclusion criteria	Exclusion criteria
Van den Bruel *et al.* 2007 [7]	GP/AP/ED	BE	4102	5.0 (0.0 to 16.9)	0.8 (0.5 to 1.1)	Children ≤16 years with acute illness max 5 days	Traumatic or neurological illness, intoxication, psychiatric or behavioral problems without somatic cause or an exacerbation of a chronic condition. No repeated inclusion of same infant within 5 days. Exclusion of physicians if the assumption of consecutive inclusion was probably violated
Roukema *et al.* 2008 [24]	ED	NL	1750	2.9 (0.1 to 15.7)	12.3 (10.8 to 13.9)	All children with fever (>38°C) at ED, without meningeal irritation	Chronic disease, Immunodeficiency
Bleeker *et al.* 2007 [25]	ED	NL	595	0.9 (0.0 to 3.0)	23.0 (19.6 to 26.4)	Children with fever T>38°C at ED, no clear focus identified after evaluation GP of history by pediatrician	Chronic disease, Immunodeficiency
Monteny *et al.* 2008 [22]	GP	NL	506	2.2 (0.3 to 5.9)	4.0 (2.3 to 5.7)	Children aged 3 months to 6 years, contacting a GP cooperative after hours with fever as the presenting symptom	Language barriers, no repeated inclusion within the previous two weeks
Brent *et al.* 2011 [23]	ED	UK	2777	3.3 (0.0 to 18.4)	5.3 (4.5 to 6.1)	All children presenting with a medical problem to the pediatric emergency-care unit whatever their age	Children who required immediate resuscitation. Comorbidity and chronic illness
Thompson *et al.* 2009 [18]	PAU	UK	700	4.6 (0.0 to 16.0)	37.7 (34.1 to 41.3)	Children aged 3 months to 16 years with suspected acute infection	Children with diseases liable to cause repeated serious bacterial infection, and infections resulting from penetrating trauma
Oostenbrink *et al.* 2004 [26]	ED	NL	593	3.7 (0.1 to 16.1)	43.8 (39.8 to 47.9)	Children aged 1 month to 16 years, meningeal signs at GP, pediatrician-referred or self-referred with neck pain	Comorbidity, ventriculoperitoneal drain

AP, ambulatory pediatric care; BE, Belgium; CI, confidence interval; ED, emergency department; GP, general practice; NL, the Netherlands; PAU, pediatric assessment unit; UK, United Kingdom.

as referral status. Temperature was recorded in all datasets (with missing data rates ranging from 0 to 18%), heart rate in five datasets (missing in 2 to 48%), capillary refill time in five (missing in 2 to 48%), respiratory rate in four (missing in 15 to 53%), and oxygen saturation in four (missing in 4 to 74%).

Validation of the FSDT [7] was possible in five datasets [18,22-25], of which four had all variables present using 'clinical sick impression' as a proxy for 'physician's gut feeling that something is wrong', and 'respiratory distress' or 'chest flaring' as a proxy for 'dyspnea' (See Additional file 2). Because the variable 'diarrhea' was missing in one dataset [25], we performed a sensitivity analysis comparing the results of the four remaining variables, as noted below.

Five datasets [18,22-25] were available for one pneumonia rule [7], developed in PC settings, with 'sick impression to clinician' as a proxy for the 'physician's gut feeling that something is wrong' and 'nasal flaring' for 'dyspnea'. A second pneumonia rule, derived in the same dataset [7], which included 'respiratory distress' and 'parental concern the illness is different' could not be validated, as the latter variable was not recorded in any of the validation datasets.

A meningitis rule, derived by Offringa *et al.* [27] for children in the ED, was validated in three datasets [7,18,26]. Because all items except 'nuchal rigidity' were present in one additional dataset [23], we performed a sensitivity analysis comparing the results of the two remaining variables, eventually excluding this dataset from the analysis, as noted below. A second meningitis rule could not be validated because the absence of its key variables in these datasets [28].

For the YOS [10], developed in secondary care, three datasets had recorded variables used in the original Yale scoring [18,22,23] (see Additional file 2). Because the YOS item 'reaction to parent stimulation' was missing in one dataset [29], we performed a sensitivity analysis comparing the results of the five remaining YOS items, as noted below. None of the datasets included sufficient variables to validate the prediction rule to identify gastroenteritis with dehydration developed by Gorelick *et al.* [30], or the prediction rule developed by Craig *et al.* [17].

The NICE guideline for feverish illness in children and the NHG alarm symptoms [15,16] were validated in four [18,22,23,26] and five [7,18,22,23,26] datasets, respectively.

Validation results
The characteristics of diagnostic accuracy, according to prevalence, are shown for all CPRs (Figure 2, Figure 3).

Low-prevalence settings
The FSDT had a sensitivity of 90% (95% CI 68 to 99%) and an NLR of 0.2 (95% CI 0.1 to 0.9) in the single LP

dataset available for validation, with false-positive test results (for example, no serious infection present) in 54% of all children examined [22]. The sensitivities of the pneumonia rule were 94% (95% CI 71 to 100%) and 92% (95% CI 86 to 96%) in two datasets, with NLRs of less than 0.2, resulting in 54% and 56% false-positive test results [22,23]. Validation of the meningitis rule in two LP datasets [7,18] resulted in sensitivities ranging from 33% (PC dataset) to 100% (secondary care dataset with a LP for meningitis) with NLRs ranging from 0.1 to 0.7. The YOS, with cut-offs of 8 and 10, provided sensitivities below 46% in one LP dataset [22], but had an NLR of greater than 0.6. The NICE 'traffic light' system with any amber or red sign present, and the NHG alarm symptoms were extremely sensitive (100%) with NLR of greater than 0.4, testing as false positive in 90 to 95% of all children in one LP dataset [22].
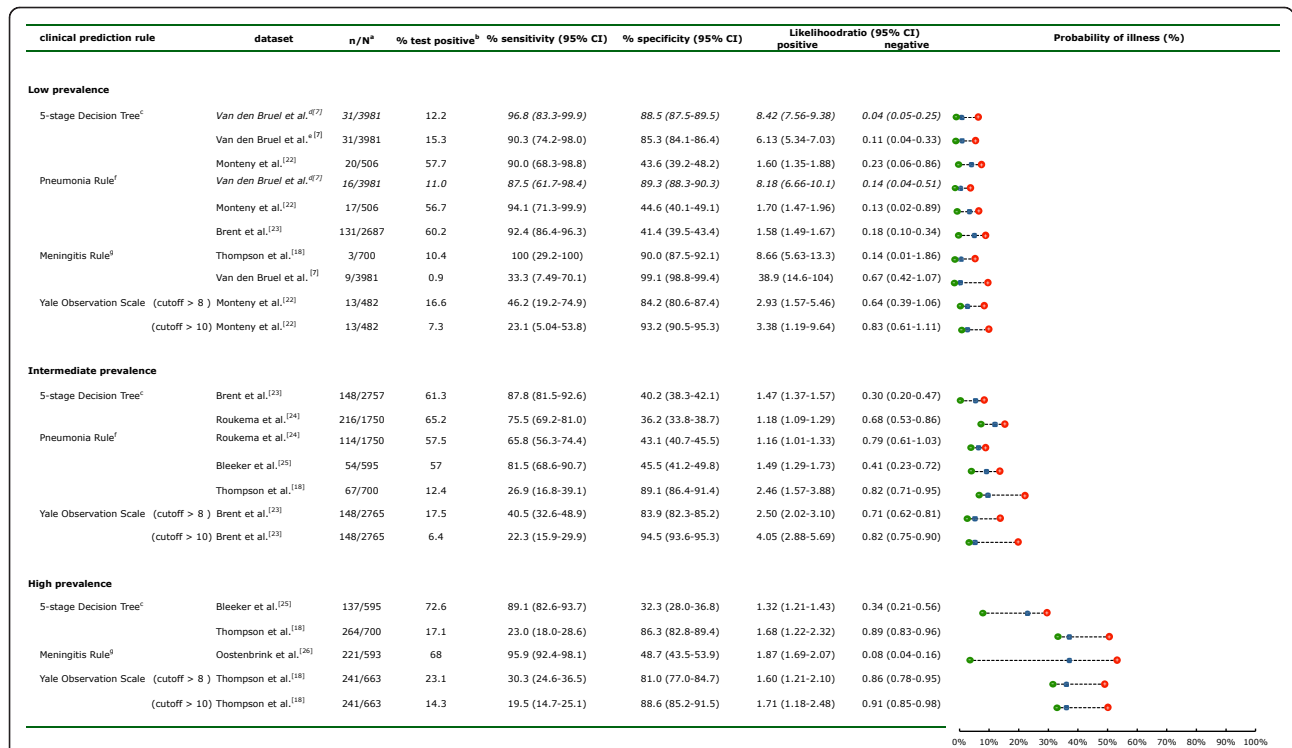
Intermediate-prevalence settings
The FSDT provided moderate sensitivities of 76% (95% CI 69 to 81%) and 88% (95% CI 82 to 93%), in two IP settings [23,24] (with NLR ranging from 0.3 to 0.7). The pneumonia rule had sensitivities ranging from 66 to 82% in two datasets [24,25] but in a third dataset [18] with the highest prevalence (11%) of pneumonia, the sensitivity was only 27% (95% CI 17 to 39%) and the NLR was 0.8 (95% CI 0.7 to 1.0). The YOS, with cut-off values of 8 and 10, provided sensitivities of less than 41% in one IP dataset [23], and had an NLR greater than 0.7. Finally, the NICE guideline and NHG alarm symptoms had high sensitivity (97 to 100%) in one IP setting [23] with NLR of less than 0.1.
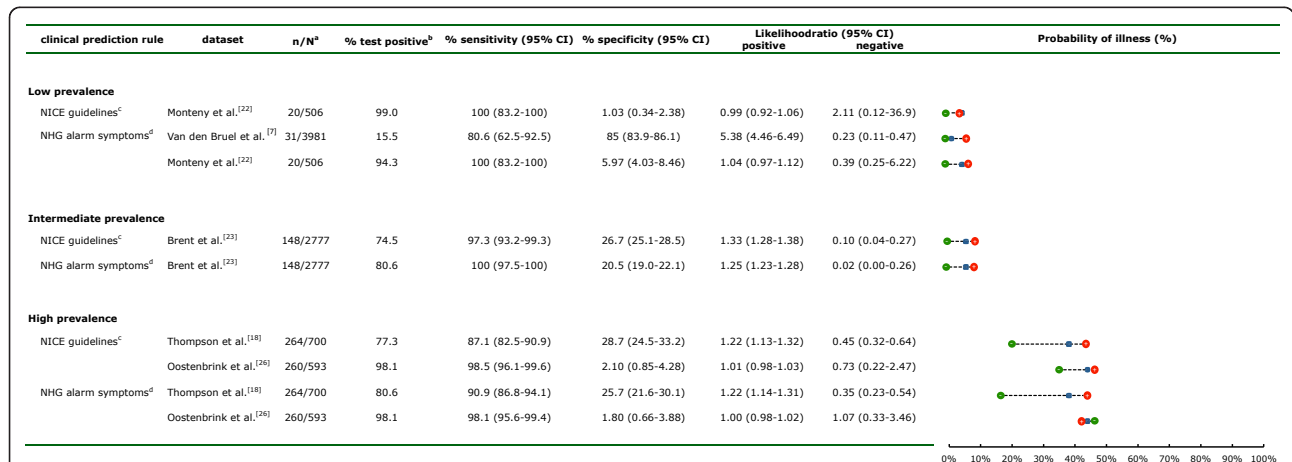
High-prevalence settings
In one HP setting [25], the FSDT had a sensitivity of 89% (95% CI 83 to 94%) with NLR of 0.4 (95% CI 0.2 to 0.6). However, sensitivity was only 23% (95% CI 18 to 29%) with NLR of 0.8 (95% CI 0.7 to 0.9) in a pediatric assessment unit [18]. In one study [26] that included children with meningeal signs identified by the referring physician, the meningitis rule showed high sensitivity, at 96% (95% CI 92 to 98%) and NLR of 0.1 (95% CI 0.04 to 0.2). The Yale score, with cut-offs of 8 and 10, provided sensitivities of less than 30% in one HP dataset [18], and NLR of 0.9. Finally, both NICE guideline and NHG alarm symptoms had sensitivities ranging from 87 to 99% in two HP datasets [18,26] with NLR greater than 0.4.

Sensitivity analyses
Comparing the 95% CIs, we found similar results for the diagnostic characteristics of the YOS and the NICE guidelines in children of all ages as well as in children for

clinical prediction rule	dataset	n/N[a]	% test positive[b]	% sensitivity (95% CI)	% specificity (95% CI)	Likelihoodratio (95% CI) positive	negative	Probability of illness (%)
Low prevalence								
5-stage Decision Tree[c]	Van den Bruel et al.[d][7]	31/3981	12.2	96.8 (83.3-99.9)	88.5 (87.5-89.5)	8.42 (7.56-9.38)	0.04 (0.05-0.25)	
	Van den Bruel et al.[e][7]	31/3981	15.3	90.3 (74.2-98.0)	85.3 (84.1-86.4)	6.13 (5.34-7.03)	0.11 (0.04-0.33)	
Pneumonia Rule[f]	Monteny et al.[22]	20/506	57.7	90.0 (68.3-98.8)	43.6 (39.2-48.2)	1.60 (1.35-1.88)	0.23 (0.06-0.86)	
	Van den Bruel et al.[d][7]	16/3981	11.0	87.5 (61.7-98.4)	89.3 (88.3-90.3)	8.18 (6.66-10.1)	0.14 (0.04-0.51)	
	Monteny et al.[22]	17/506	56.7	94.1 (71.3-99.9)	44.6 (40.1-49.1)	1.70 (1.47-1.96)	0.13 (0.02-0.89)	
	Brent et al.[23]	131/2687	60.2	92.4 (86.4-96.3)	41.4 (39.5-43.4)	1.58 (1.49-1.67)	0.18 (0.10-0.34)	
Meningitis Rule[g]	Thompson et al.[18]	3/700	10.4	100 (29.2-100)	90.0 (87.5-92.1)	8.66 (5.63-13.3)	0.14 (0.01-1.86)	
	Van den Bruel et al.[7]	9/3981	0.9	33.3 (7.49-70.1)	99.1 (98.8-99.4)	38.9 (14.6-104)	0.67 (0.42-1.07)	
Yale Observation Scale (cutoff > 8)	Monteny et al.[22]	13/482	16.6	46.2 (19.2-74.9)	84.2 (80.6-87.4)	2.93 (1.57-5.46)	0.64 (0.39-1.06)	
(cutoff > 10)	Monteny et al.[22]	13/482	7.3	23.1 (5.04-53.8)	93.2 (90.5-95.3)	3.38 (1.19-9.64)	0.83 (0.61-1.11)	
Intermediate prevalence								
5-stage Decision Tree[c]	Brent et al.[23]	148/2757	61.3	87.8 (81.5-92.6)	40.2 (38.3-42.1)	1.47 (1.37-1.57)	0.30 (0.20-0.47)	
Pneumonia Rule[f]	Roukema et al.[24]	216/1750	65.2	75.5 (69.2-81.0)	36.2 (33.8-38.7)	1.18 (1.09-1.29)	0.68 (0.53-0.86)	
	Roukema et al.[24]	114/1750	57.5	65.8 (56.3-74.4)	43.1 (40.7-45.5)	1.16 (1.01-1.33)	0.79 (0.61-1.03)	
	Bleeker et al.[25]	54/595	57	81.5 (68.6-90.7)	45.5 (41.2-49.8)	1.49 (1.29-1.73)	0.41 (0.23-0.72)	
	Thompson et al.[18]	67/700	12.4	26.9 (16.8-39.1)	89.1 (86.4-91.4)	2.46 (1.57-3.88)	0.82 (0.71-0.95)	
Yale Observation Scale (cutoff > 8)	Brent et al.[23]	148/2765	17.5	40.5 (32.6-48.9)	83.9 (82.3-85.2)	2.50 (2.02-3.10)	0.71 (0.62-0.81)	
(cutoff > 10)	Brent et al.[23]	148/2765	6.4	22.3 (15.9-29.9)	94.5 (93.6-95.3)	4.05 (2.88-5.69)	0.82 (0.75-0.90)	
High prevalence								
5-stage Decision Tree[c]	Bleeker et al.[25]	137/595	72.6	89.1 (82.6-93.7)	32.3 (28.0-36.8)	1.32 (1.21-1.43)	0.34 (0.21-0.56)	
	Thompson et al.[18]	264/700	17.1	23.0 (18.0-28.6)	86.3 (82.8-89.4)	1.68 (1.22-2.32)	0.89 (0.83-0.96)	
Meningitis Rule[g]	Oostenbrink et al.[26]	221/593	68	95.9 (92.4-98.1)	48.7 (43.5-53.9)	1.87 (1.69-2.07)	0.08 (0.04-0.16)	
Yale Observation Scale (cutoff > 8)	Thompson et al.[18]	241/663	23.1	30.3 (24.6-36.5)	81.0 (77.0-84.7)	1.60 (1.21-2.10)	0.86 (0.78-0.95)	
(cutoff > 10)	Thompson et al.[18]	241/663	14.3	19.5 (14.7-25.1)	88.6 (85.2-91.5)	1.71 (1.18-2.48)	0.91 (0.85-0.98)	

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Figure 2 Results of external validation of clinical prediction rules (CPRs) to rule in or rule out serious infection. [a]Number of cases (n) out of the total population of all children (N). [b]Percentage testing positive in all included children. [c]If yes to any of five sequential questions: 1) clinical instinct that something is wrong, 2) dyspnea, 3) temperature greater than 39.5°C, 4) diarrhea, 5) age 15 to 29 months; [d]Derivation study (italic). [e]'clinical instinct that something is wrong' replaced by 'clinical impression'. [f]If yes to any of the following: 1) shortness of breath, 2) clinicians concern. [g]If yes to any of the following: 1) petechiae, 2) nuchal rigidity, 3) coma; probability of illness (in percentage) before testing (blue dot), after a positive test result (red dot with plus to sign) and after a negative test result (green dot with minus to sign).

clinical prediction rule	dataset	n/N[a]	% test positive[b]	% sensitivity (95% CI)	% specificity (95% CI)	Likelihoodratio (95% CI) positive	negative	Probability of illness (%)
Low prevalence								
NICE guidelines[c]	Monteny et al.[22]	20/506	99.0	100 (83.2-100)	1.03 (0.34-2.38)	0.99 (0.92-1.06)	2.11 (0.12-36.9)	
NHG alarm symptoms[d]	Van den Bruel et al.[7]	31/3981	15.5	80.6 (62.5-92.5)	85 (83.9-86.1)	5.38 (4.46-6.49)	0.23 (0.11-0.47)	
	Monteny et al.[22]	20/506	94.3	100 (83.2-100)	5.97 (4.03-8.46)	1.04 (0.97-1.12)	0.39 (0.25-6.22)	
Intermediate prevalence								
NICE guidelines[c]	Brent et al.[23]	148/2777	74.5	97.3 (93.2-99.3)	26.7 (25.1-28.5)	1.33 (1.28-1.38)	0.10 (0.04-0.27)	
NHG alarm symptoms[d]	Brent et al.[23]	148/2777	80.6	100 (97.5-100)	20.5 (19.0-22.1)	1.25 (1.23-1.28)	0.02 (0.00-0.26)	
High prevalence								
NICE guidelines[c]	Thompson et al.[18]	264/700	77.3	87.1 (82.5-90.9)	28.7 (24.5-33.2)	1.22 (1.13-1.32)	0.45 (0.32-0.64)	
	Oostenbrink et al.[26]	260/593	98.1	98.5 (96.1-99.6)	2.10 (0.85-4.28)	1.01 (0.98-1.03)	0.73 (0.22-2.47)	
NHG alarm symptoms[d]	Thompson et al.[18]	264/700	80.6	90.9 (86.8-94.1)	25.7 (21.6-30.1)	1.22 (1.14-1.31)	0.35 (0.23-0.54)	
	Oostenbrink et al.[26]	260/593	98.1	98.1 (95.6-99.4)	1.80 (0.66-3.88)	1.00 (0.98-1.02)	1.07 (0.33-3.46)	

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Figure 3 Results of external validation of the evidence-based clinical guidelines for management of fever. [a]Number of cases (n) out of the total population of all children (N). [b]Percentage testing positive in all included children. [c]'Traffic light' system of clinical features that are designed to be used to assess the risk of serious infection, and to provide clinical guidance for actions needed according to these categories. [d]Alarm symptoms at clinical examination: seriously ill impression, reduced consciousness, persistent vomiting, petechiae, tachypnea and/or dyspnea, reduced peripheral circulation, pallor, and signs of meningeal irritation; probability of illness (in percentage) before testing (blue dot), after a positive test result (red dot with + sign) and after a negative test result (green dot with - sign).

whom the rules were originally designed (3 to 35 months and up to 5 years, respectively) (see Additional file 4).

Comparing the results of the datasets in which the complete prediction rule could be validated with those of the datasets with one or two missing variables (five items of the YOS, four items of the FSDT, and six items of the NHG alarm symptoms), all diagnostic characteristics were found to be similar through comparison of the 95% CIs (see Additional file 4).

By contrast, dropping 'nuchal rigidity' from the meningitis rule resulted in a lower sensitivity (67% (95% CI 9 to 99%) versus 100% (95% CI 29 to 100%) when all three variables were considered) in one dataset [18], eliminating one additional dataset, which had only two out of three original variables available, for further use in the validation [23].

Discussion
Main findings
None of the CPRs examined in this study provided perfect diagnostic accuracy. The best performing CPR for ruling out serious infection in an LP setting was the FSDT, which uses the physician's gut feeling, the patient's age and temperature, and presence of dyspnea and diarrhea [7]. Sensitivity was lower than that reported in the original study, possibly explained by our use of 'clinical impression' as a proxy for 'physician's gut feeling' which has been reported to be of lower diagnostic value [3].

Both the NICE guideline and the NHG alarm symptoms high sensitivity in both LP and IP settings, suggesting possible clinical value for ruling out serious infections in children presenting in these settings. However, large numbers of children were flagged as potentially having a serious infection. If the prediction rules were to be used in clinical practice, additional clinical assessment, additional testing, or review at a later stage would be necessary to avoid inappropriate referrals or hospital admissions.

For the well-known YOS, all sensitivities were low, which is similar to the results of a previously reported pooled sensitivity based on the meta-analysis of seven studies [3].

Other disease-specific rules (pneumonia and meningitis) had acceptable sensitivities only in the LP settings, indicating value as rule-out tests. However, the percentage of false positives was too high in all datasets, apart from one IP dataset, probably due to the higher prevalence of pneumonia in this dataset [18].

Limitations
Despite the large number of datasets available, we were able to validate only four of the eight prediction rules plus both guidelines. The methodological challenges encountered in performing these retrospective validations in prospectively collected datasets limit the translation into clinical practice. Performance of prediction rules was generally lower than in their original derivation studies. One possible explanation for this is the approximations that we used for variables measured and recorded in different ways (and different languages).

To avoid potential bias from validating in datasets that were missing variables, a sensitivity analyses was performed and, if findings were robust throughout the different validation datasets, subsequent validation was deemed suitable. In addition to variation in recorded variables, multiple other sources of heterogeneity were found in the included databases, including differences in setting, inclusion criteria, immunization schedules, and definition of serious infection.

Strengths
Although the limitations may be substantial, this is the first study to externally validate existing CPRs in different types of clinical settings. We used individual patient data from a total of seven existing datasets comprising 11,023 children presenting to PC or EDs in three European countries to retrospectively validate existing prediction rules and national evidence-based guidelines. Previously, only a single prediction rule had been prospectively validated in external datasets [11-14]. Our study therefore presents the first robust attempt to simultaneously validate multiple current prediction rules and evidence-based guidelines for management of one of the most common clinical conditions in AC settings. We anticipate that our findings will be applicable to guideline developers worldwide.

Comparison with other studies
The YOS was initially developed to identify serious illness in febrile children aged 3 to 36 months, but was subsequently discarded based on three prospective validation studies (of which only one was carried out in the intended age group) [11,13,14]. The rule was also used to stratify patients in five studies evaluating inflammatory markers (such as procalcitonin and C-reactive protein), with discouraging results [31-35]. Bang *et al.* reported a slightly better performance of the YOS in predicting bacteremia in febrile children in an HP study (28%), which does not apply to most AC settings [12]. Although the YOS was not useful for ruling out a serious infection in our analysis, a score of greater than 10 (with a combination of the presence of abnormal color or hydration status, failure to respond to parents, different cry, and abnormal sleepiness) did slightly increase the likelihood of a serious infection in these datasets.

Clinical implications
With decreasing incidence of serious infections, clinicians will increasingly rely on CPRs in practice, particularly in high-volume triage settings. In these settings,

'generic' rules, which apply to all serious infections, are more useful than disease-specific rules. Particularly in settings where diagnosis of serious illness in children is essential (for example, PC), the FSDT, the NICE guidelines, and the NHG alarm symptoms may be used to rule out serious infections in a large proportion of children. We suggest that the FSDT, mainly consisting of the child's breathing status and temperature and the clinician's gut feeling that something is wrong, should be used for assessment of every acutely ill child. The meningitis rule, with absence of nuchal rigidity, petechiae, and coma, indicate that meningitis is highly unlikely in LP settings.

Clinicians should be aware that none of the CPRs provide perfect discrimination, and it is perhaps unrealistic to expect such rules to provide this. Residual uncertainty may be further improved by conducting more detailed clinical assessments, repeating the assessment after some time, using additional testing (for example, urine or blood tests), and in most cases, providing an appropriate safety netting advice for children sent home detailing instructions on when to seek further care [36].

Research implications

Most CPRs never undergo further validation or are implemented, perhaps inappropriately, with insufficient external validation [37,38]. Indeed very few CPRs for the identification of children with serious infection have undergone either extensive validation or formal impact analysis, limiting the ability to truly evaluate their performance and to balance benefits and harms [19,39]. In general, CPRs perform worse when validated in new populations [40].

Our study presents the first multiple external validation of CPRs in this common clinical area, and identifies which of them offer the best diagnostic accuracy in different types of clinical settings. This illustrates the clear need to perform extensive prospective validation and impact analysis of CPRs prior to clinical implementation [39,41]. The FSDT and the NICE guidelines for assessment of feverish children are potential candidates for future prospective validation studies examining their performance in new prospectively collected data on similar populations.

We recognize the previously identified major mismatch [3], between the clinical settings where the majority of children with acute infections seek help (that is, PC), and the number of studies performed in that setting (two studies) (Table 2). There is a pressing need for more studies conducted in PC or in LP ED settings to validate CPRs for serious infection, or the need for hospital referral/admission. Given the relative infrequency of serious infections, such studies need to include large cohorts of children [7,8]. CPRs are mostly designed to rule out serious infections, often at the expense of moderate to low ability for inclusion. As no rule is perfect at ruling out infection,

research on the most effective content and methods of delivery with appropriate safety netting advice in PC and EDs is essential [8,36,42]. Adding newer tests such as point-of-care inflammatory markers may improve the diagnostic value of these rules, but the performance of these markers in non-referred populations has to be tested [43].

Conclusions

None of the CPRs examined in this study provided perfect diagnostic accuracy. In LP settings (for example, PC) or IP settings, prediction rules, such as the FSDT and evidence-based guidelines (NICE guideline and the NHG alarm symptoms) had high sensitivity, providing promising rule-out value for serious infections in these datasets, although all seemed to leave residual uncertainty. Additional clinical assessment or testing such as point-of-care inflammatory markers may be needed to increase clinical certainty. None of the prediction rules identified seemed to be valuable for HP settings (for example, EDs).

Additional material

Additional file 1: Details of the clinical prediction rules identified in the systematic review. CRT, capillary refill time; RR, respiratory rate; Temp, temperature.

Additional file 2: variables and proxies used for validation of clinical prediction rules. N = Number of children in dataset; % n/N = Percentage of cases (n) out of all children (N) used for the external validation analysis; green font indicate original variable, red font,variable not recorded, blue font, proxy variable. [a]Derivation study (italic). [b]'Clinical sick impression' used as proxy for 'physician's gut feeling that something is wrong', [c]'Respiratory distress' used as proxy for 'dyspnoea'. [d]'Chest flaring' used as proxy for 'dyspnoea'. [e]'Meningeal irritation' used as proxy for 'nuchal rigidity'. [f]'Unconsciousness' used as proxy for 'coma'.

Additional file 3: Variables and proxies used for fever guidelines validation. N. number of children in dataset; % n/N. percentage of cases (n) out of all children (N) used for the external validation analysis; green font. original variable

Additional file 4: Sensitivity analyses. CI, confidence interval; underlined, 95% CIs not comparable.

Authors' contributions

JV undertook the translation, the synopsis, the re-coding and the data checking and the results of each step were discussed with all study authors (AVdB, MB, RO, HM, MT, ML), contributing data to the current study. JV undertook the external validation analysis and drafted the report. AVdB, MT, RS, BA, RO, HM, MB, ML, DM, and FB conceived the analyses, co-drafted the report, and commented on it. All authors have read and approved the final manuscript.

Author details
[1]Department of General Practice, KU Leuven, Kapucijnenvoer 33 blok J, 3000 Leuven, Belgium. [2]Department of Primary Care Health Sciences, University of Oxford, New Radcliffe House, Woodstock Road, Oxford, OX2 6GG, UK. [3]Erasmus MC - Sophia's Children Hospital, Dr Molewaterplein 60, 3015 GJ Rotterdam, The Netherlands. [4]Department of General Practice, University Medical Center Groningen, Hanze plein 1 Box 30001 9700RB Groningen, The Netherlands. [5]Department of General and Adolescent Paediatrics, University College London, Institute of Child Health, London, UK. [6]Research Institute Caphri, Maastricht University, PB 313, NI 6200 MD, Maastricht, The Netherlands.

References
1. Hay AD, Heron J, Ness A: **The prevalence of symptoms and consultations in pre-school children in the Avon Longitudinal Study of Parents and Children (ALSPAC): a prospective cohort study.** *Fam Pract* 2005, **22**:367-374.
2. Van den Bruel A, Bartholomeeusen S, Aertgeerts B, Truyers C, Buntinx F: **Serious infections in children: an incidence study in family practice.** *BMC Fam Pract* 2006, **7**:23-23.
3. 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**:834-845.
4. Strang JR, Pugh EJ: **Meningococcal infections: reducing the case fatality rate by giving penicillin before admission to hospital.** *BMJ* 1992, **305**:141-143.
5. Koomen I, Grobbee DE, Roord JJ, Donders R, Jennekens-Schinkel A, van Furth AM: **Hearing loss at school age in survivors of bacterial meningitis: assessment, incidence, and prediction.** *Pediatrics* 2003, **112**:1049-1053.
6. CEMACH: **Why children die: a pilot study 2006; England (South West, North East and West Midlands), Wales and Northern Ireland.** *CEMACH* 2008.
7. 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**:538-546.
8. Buntinx F, Mant D, Van den Bruel A, Donner-Banzhof N, Dinant GJ: **Dealing with low-incidence serious diseases in general practice.** *Br J Gen Pract* 2011, **61**:43-46.
9. Oostenbrink R, Thompson M, Steyerberg EW: **Barriers to translating diagnostic research in febrilechildren to clinical practice: a systematic review.** *Arch Dis Child* 2012.
10. McCarthy PL, Sharpe MR, Spiesel SZ, Dolan TF, Forsyth BW, DeWitt TG, Fink HD, Baron MA, Cicchetti DV: **Observation scales to identify serious illness in febrile children.** *Pediatrics* 1982, **70**:802-809.
11. Baker MD, Avner JR, Bell LM: **Failure of infant observation scales in detecting serious illness in febrile, 4- to 8-week-old infants.** *Pediatrics* 1990, **85**:1040-1043.
12. Bang A, Chaturvedi P: **Yale Observation Scale for prediction of bacteremia in febrile children.** *Indian J Pediatr* 2009, **76**:599-604.
13. Hsiao AL, Chen L, Baker MD: **Incidence and predictors of serious bacterial infections among 57- to 180-day-old infants.** *Pediatrics* 2006, **117**:1695-1701.
14. Teach SJ, Fleisher GR: **Efficacy of an observation scale in detecting bacteremia in febrile children three to thirty-six months of age, treated as outpatients. Occult Bacteremia Study Group.** *J Pediatr* 1995, **126**:877-881.
15. NICE: **National Institute for Clinical Excellence: Feversh illness in children - assessment and initial management in children younger than 5 years.** *London: National Institute for Health and Clinical Excellence, 2007* 2007, [cited June 29, 2007].
16. Berger MY BL, Albeda FW, Dijkstra RH, Graafmans TA, Van der Laan JR, Lemmen WH, Oteman N: **NHG-Standaard Kinderen met koorts - Tweede Herziening.** *Huisarts Wet* 2008, **51**:287-296.
17. Craig JC, Williams GJ, Jones M, Codarini M, Macaskill P, Hayen A, Irwig L, Fitzgerald DA, Isaacs D, McCaskill M: **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**:c1594.
18. Thompson M, Coad N, Harnden A, Mayon-White R, Perera R, Mant D: **How well do vital signs identify children with serious infections in paediatric emergency care?** *Arch Dis Child* 2009, **94**:888-893.
19. Oostenbrink R, Thompson M, Steyerberg EW: **Barriers to translating diagnostic research in febrile children to clinical practice: a systematic review.** *Arch Dis Child* 2011.
20. 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**:703-707.
21. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y: **Analysing and presenting results.** In *Chapter 10. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Volume Version 1.0.* Edited by: Deeks J, Bossuyt PM, Gatsonis C. The Cochrane Collaboration; 2010, Available from: [http://srdta.cochrane.org/] (released December 23rd, 2010).
22. Monteny M, Berger MY, van der Wouden JC, Broekman BJ, Koes BW: **Triage of febrile children at a GP cooperative: determinants of a consultation.** *Br J Gen Pract* 2008, **58**:242-247.
23. Brent AJ, Lakhanpaul M, Thompson M, Collier J, Ray S, Ninis N, Levin M, Macfaul R: **Risk score to stratify children with suspected serious bacterial infection: observational cohort study.** *Arch Dis Child* 2011, **96**:6.
24. 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**:107-113.
25. Bleeker SE, Derksen-Lubsen G, Grobbee DE, Donders AR, Moons KG, Moll HA: **Validating and updating a prediction rule for serious bacterial infection in patients with fever without source.** *Acta paediatrica* 2007, **96**:100-104.
26. Oostenbrink R, Moons KG, Derksen-Lubsen AG, Grobbee DE, Moll HA: **A diagnostic decision rule for management of children with meningeal signs.** *Eur J Epidemiol* 2004, **19**:109-116.

27. Offringa M, Beishuizen A, Derksen-Lubsen G, Lubsen J: **Seizures and fever: can we rule out meningitis on clinical grounds alone?** *Clin Pediatr (Phila)* 1992, **31**:514-522.
28. Joffe A, McCormick M, DeAngelis C: **Which children with febrile seizures need lumbar puncture? A decision analysis approach.** *Am J Dis Child* 1983, **137**:1153-1156.
29. Brent AJ, Lakhanpaul M, Thompson M, Collier J, Ray S, Ninis N, Levin M, Macfaul R: **Risk score to stratify children with suspected serious bacterial infection: observational cohort study.** *Arch Dis Child* 2011, **96**:361-367.
30. Gorelick MH, Shaw KN, Murphy KO: **Validity and reliability of clinical signs in the diagnosis of dehydration in children.** *Pediatrics* 1997, **99**:E6.
31. Andreola B, Bressan S, Callegaro S, Liverani A, Plebani M, Da Dalt L: **Procalcitonin and C-reactive protein as diagnostic markers of severe bacterial infections in febrile infants and children in the emergency department.** *Pediatr Infect Dis J* 2007, **26**:672-677.
32. Galetto-Lacour A, Zamora SA, Gervaix A: **Bedside procalcitonin and C-reactive protein tests in children with fever without localizing signs of infection seen in a referral center.** *Pediatrics* 2003, **112**:1054-1060.
33. Lacour AG, Gervaix A, Zamora SA, Vadas L, Lombard PR, Dayer JM, Suter S: **Procalcitonin, IL-6, IL-8, IL-1 receptor antagonist and C-reactive protein as identificators of serious bacterial infections in children with fever without localising signs.** *Eur J Pediatr* 2001, **160**:95-100.
34. Pulliam PN, Attia MW, Cronan KM: **C-reactive protein in febrile children 1 to 36 months of age with clinically undetectable serious bacterial infection.** *Pediatrics* 2001, **108**:1275-1279.
35. Thayyil S, Shenoy M, Hamaluba M, Gupta A, Frater J, Verber IG: **Is procalcitonin useful in early diagnosis of serious bacterial infections in children?** *Acta paediatrica* 2005, **94**:155-158.
36. Almond S, Mant D, Thompson M: **Diagnostic safety-netting.** *Br J Gen Pract* 2009, **59**:872-874, discussion 874.
37. Reilly BM, Evans AT: **Translating clinical research into clinical practice: impact of using prediction rules to make decisions.** *Ann Intern Med* 2006, **144**:201-209.
38. Maguire JL, Kulik DM, Laupacis A, Kuppermann N, Uleryk EM, Parkin PC: **Clinical prediction rules for children: a systematic review.** *Pediatrics* 2011.
39. Wallace E, Smith SM, Perera-Salazar R, Vaucher P, McCowan C, Collins G, Verbakel J, Lakhanpaul M, Fahey T: **Framework for the impact analysis and implementation of clinical prediction rules (CPRs).** *BMC Med Inform Decis Mak* 2011, **11**:62.
40. Justice AC, Covinsky KE, Berlin JA: **Assessing the generalizability of prognostic information.** *Ann Intern Med* 1999, **130**:515-524.
41. Van den Bruel A, Cleemput I, Aertgeerts B, Ramaekers D, Buntinx F: **The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed.** *Journal of clinical epidemiology* 2007, **60**:1116-1122.
42. Maguire S, Ranmal R, Komulainen S, Pearse S, Maconochie I, Lakhanpaul M, Davies F, Kai J, Stephenson T: **Which urgent care services do febrile children use and why?** *Arch Dis Child* 2011, **96**:810-816.
43. Van den Bruel A, Thompson MJ, Haj-Hassan T, Stevens R, Moll H, Lakhanpaul M, Mant D: **Diagnostic value of laboratory tests in identifying serious infections in febrile children: systematic review.** *BMJ* 2011, **342**: d3082.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1741-7015/11/10/prepub