Diagnostic value of laboratory tests in identifying serious infections in febrile children: systematic review

Ann Van den Bruel, academic clinical lecturer,1 Matthew J Thompson, associate professor,1,2 Tanya Haj-Hassan, medical student,3 Richard Stevens, senior statistician,1 Henriette Moll, professor,4 Monica Lakhanpaul, senior lecturer,2 David Mant, emeritus professor1

ABSTRACT
Objective To collate all available evidence on the diagnostic value of laboratory tests for the diagnosis of serious infections in febrile children in ambulatory settings.

Design Systematic review.

Data sources Electronic databases, reference tracking, and consultation with experts.

Study selection Studies were selected on six criteria: design (studies of diagnostic accuracy or deriving prediction rules), participants (otherwise healthy children and adolescents aged 1 month to 18 years), setting (first contact ambulatory care), outcome (serious infection), features assessed (in first contact care), and data reported (sufficient to construct a 2×2 table).

Data extraction Quality assessment was based on the QUADAS (Quality assessment tool of diagnostic accuracy studies) criteria. Meta-analyses were done using the bivariate random effects method and hierarchical summary receiver operating characteristic curves for studies with multiple thresholds.

Data synthesis None of the 14 studies identified were of high methodological quality and all were carried out in an emergency department or paediatric assessment unit. The prevalence of serious infections ranged from 4.5% to 29.3%. Tests were carried out for C reactive protein (five studies), procalcitonin (three), erythrocyte sedimentation rate (one), interleukins (two), white blood cell count (seven), absolute neutrophil count (two), band count (three), and left shift (one). The tests providing most diagnostic value were C reactive protein and procalcitonin. Bivariate random effects meta-analysis (five studies, 1379 children) for C reactive protein yielded a pooled positive likelihood ratio of 3.15 (95% confidence interval 2.67 to 3.71) and a pooled negative likelihood ratio of 0.07 (0.02 to 0.27). To rule in serious infection (positive likelihood ratio 4.92 (3.26 to 7.43) and a negative likelihood ratio of 0.07 (0.02 to 0.27).

Conclusion Measuring inflammatory markers in an emergency department setting can be diagnostically useful, but clinicians should apply different cut-off values depending on whether they are trying to rule in or rule out serious infection. Measuring white blood cell count is less useful for ruling in serious infection and not useful for ruling out serious infection. More rigorous studies are needed, including studies in primary care, to assess the value of laboratory tests alongside clinical diagnostic measurements, including vital signs.

INTRODUCTION
We recently published a systematic review on the diagnostic value of presenting clinical features in identifying serious infection in children.1 This review identified several important red flags; it also confirmed that consideration of symptoms and signs alone often results in residual diagnostic uncertainty—with the risk of serious infection being too high to ignore yet too low to justify hospital admission. The diagnostic uncertainty that clinicians are left with after clinical assessment was confirmed in a recent, large cohort study, where even a complex clinical decision rule involving 28 clinical features could not provide perfect discrimination.

In situations with a significant risk of rapid progression of illness, or where further refining of the risk estimate could either rule out serious infection or influence a decision to admit to hospital or treat with antibiotics,3 clinicians often try to increase diagnostic certainty by measuring the white blood cell count or blood levels of inflammatory markers. Although in a hospital setting these tests are normally carried out in a laboratory, many are now available as point of care tests, which give an immediate result and can be used in ambulatory care settings.
Earlier systematic reviews have assessed the value of one or two laboratory tests in children only or in children and adults for the diagnosis of various outcomes, such as distinguishing viral pneumonia from bacterial pneumonia or parenchymal involvement in children admitted with a urinary tract infection. We reviewed the diagnostic value of all possible blood tests for ruling in and ruling out serious infection in children in ambulatory settings and their added value after clinical signs and symptoms, using the same rigorous methods as in our previous review.

METHODOLOGY

The methods of our review have been published in detail elsewhere. Briefly, we searched the literature electronically in Medline, Embase, DARE, and CINAHL, using search terms relating to serious infections, children, clinical and laboratory tests, and ambulatory care (see web extra annex 1). In addition, a snowballing strategy included checking the reference lists of included studies, systematic reviews, and relevant National Institute for Health and Clinical Excellence guidelines, and consultation with experts.

Two independent reviewers (AVDB and TH-H) carried out the selection. A third independent reviewer (MJT) resolved discrepancies. Studies were selected if they assessed diagnostic accuracy of clinical features or laboratory tests for ruling in and ruling out serious infection in previously well children and adolescents between 1 month and 18 years of age presenting to ambulatory care (defined as general or family practice, paediatric outpatient clinics, paediatric assessment units, or emergency departments) in Europe, Canada, the United States, Australia, New Zealand, and Japan.

Serious infections were defined as sepsis (including bacteremia), meningitis, pneumonia, osteomyelitis, cellulitis, gastroenteritis with dehydration, complicated urinary tract infection [positive urine culture result and systemic effects such as fever], and viral respiratory tract infections complicated by hypoxia (for example, bronchiolitis). Studies restricted to one specific serious infection (bacteremia or meningitis) were not selected for this paper but will be reported separately.

We selected studies if they reported sufficient data to allow construction of $2 \times 2$ tables. No language restrictions were applied.

Quality assessment

Two independent reviewers (AVDB and TH-H) used the QUADAS instrument to assess the quality of selected articles. Any disagreements were resolved by discussion and consensus, if necessary after contacting the authors for clarification. We used spectrum bias and validity of reference standards as exclusion criteria.

Studies selected for analysis were given an A, B, C, or D rating. We rated studies fulfilling all QUADAS criteria as A; studies with no or unclear total verification with the reference standard or with interpretation of the index feature unblinded to the results of the reference standard as D; and studies without an independent reference standard, with interpretation of the reference standard unblinded to the results of the index feature, or with an unduly long period between recording of the index feature and outcome as C. All other studies were rated B. If data were insufficient to be confident that a criterion had been met, we assessed the criterion as not being met.

Data extraction and analysis

Data were extracted by one reviewer (AVDB) and checked by a second reviewer (TH-H); $2 \times 2$ tables were constructed based on information in the article or from the authors. We then calculated the likelihood ratios for the presence or absence of each feature. Unlike sensitivity and specificity, likelihood ratios make explicit the impact of the test result on the probability of the disease. It is easiest for clinicians to think of a positive result with a likelihood ratio of 2 as making it twice as likely that the patient has the disease. However, this is imprecise as the likelihood ratio applies to the change in odds rather than probability of disease—that is, the likelihood ratio $\times$ pre-test odds $=$ post-test odds. So to calculate the precise impact of the test result on disease probability, it is necessary first to convert the pre-test probability to odds (pre-test odds $=$ pre-test probability$/ (1 −$ pre-test probability)) and then after multiplying by the likelihood ratio convert the odds back to probability (post-test probability $=$ post-test odds$/ (1+post-test odds))

![Flow of papers through study](Figure 1)
<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Setting, country</th>
<th>No of participants</th>
<th>% serious infection</th>
<th>Quality rating</th>
<th>Age range</th>
<th>Inclusion criteria</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andreola 2007</td>
<td>Prospective, cross sectional, consecutive</td>
<td>Emergency department, Italy</td>
<td>408</td>
<td>23.0</td>
<td>C</td>
<td>&lt;3 years</td>
<td>Fever of uncertain source and increased risk of serious bacterial infection; namely, all infants aged 7 days to 3 months with rectal temperature ≥38°C and children aged 3-36 months with ill/toxic appearance or with rectal temperature &gt;39.5°C</td>
<td>Bacteraemia, sepsis, acute pyelonephritis, lobar pneumonia, bacterial meningitis, bone or joint infections</td>
</tr>
<tr>
<td>Baker 1993</td>
<td>Prospective, cross sectional, consecutive</td>
<td>Emergency department, USA</td>
<td>747</td>
<td>8.7</td>
<td>C</td>
<td>29-56 days</td>
<td>Temperature (rectal) ≥38.2°C and immunocompetent</td>
<td>Bacteraemia, bacterial meningitis, bacterial gastroenteritis, urinary tract infection, pneumonia, aseptic meningitis, cellulitis, abscess</td>
</tr>
<tr>
<td>Baker 1999</td>
<td>Prospective, cross sectional, consecutive</td>
<td>Emergency department, USA</td>
<td>422</td>
<td>10.2</td>
<td>C</td>
<td>29-60 days</td>
<td>Temperature (rectal) ≥38.0°C and immunocompetent</td>
<td>Bacteraemia, bacterial meningitis, bacterial gastroenteritis, urinary tract infection, pneumonia, aseptic meningitis, cellulitis, abscess, bone or joint infection</td>
</tr>
<tr>
<td>Berger 1996</td>
<td>Prospective, cross sectional, consecutive</td>
<td>Emergency department, Netherlands</td>
<td>138</td>
<td>23.9</td>
<td>B</td>
<td>2 weeks -1 year</td>
<td>Temperature (rectal) ≥38.0°C measured on the ward</td>
<td>Pneumonia, urinary tract infection, bacteraemia, meningitis, cellulitis, septic arthritis, osteomyelitis, otitis media, bacterial gastroenteritis</td>
</tr>
<tr>
<td>Bleeker 2007</td>
<td>Prospective, cross sectional, consecutive</td>
<td>Emergency department, Netherlands</td>
<td>381</td>
<td>26.0</td>
<td>D</td>
<td>1-36 months</td>
<td>Referred to emergency department for fever without source—temperature ≥38°C for which no clear focus could be identified after evaluation by general practitioner or after history taking by paediatrician</td>
<td>Bacterial meningitis, sepsis or bacteraemia, urinary tract infection, pneumonia, bacterial gastroenteritis, osteomyelitis, ethmoiditis</td>
</tr>
<tr>
<td>Bonadio 1993</td>
<td>Prospective, cross sectional, consecutive</td>
<td>Emergency department, USA</td>
<td>534</td>
<td>4.5</td>
<td>D</td>
<td>4-8 weeks</td>
<td>Temperature (rectal) ≥38.0°C at triage, previously healthy</td>
<td>Bacterial meningitis, bacteraemia, urinary tract infection, salmonella enteritis, osteomyelitis, septic arthritis</td>
</tr>
<tr>
<td>Galetto-Lacour 2001</td>
<td>Prospective, cross sectional</td>
<td>Emergency department, Switzerland</td>
<td>124</td>
<td>22.6</td>
<td>D</td>
<td>7 days -36 months</td>
<td>Temperature (rectal) ≥38.0°C and no localising signs of infection from history or physical examination</td>
<td>Bacteraemia, pyelonephritis, lobar pneumonia, meningitis, osteoarthritis</td>
</tr>
<tr>
<td>Galetto-Lacour 2003</td>
<td>Prospective, cross sectional</td>
<td>Emergency department, Switzerland</td>
<td>99</td>
<td>29.3</td>
<td>D</td>
<td>7 days -36 months</td>
<td>Temperature (rectal) ≥38.0°C and no localising signs of infection in history or at physical examination</td>
<td>Bacteraemia, pyelonephritis, lobar pneumonia, meningitis, osteoarthritis</td>
</tr>
<tr>
<td>Galetto-Lacour 2008</td>
<td>Prospective, cross sectional</td>
<td>Emergency department, Switzerland</td>
<td>202</td>
<td>26.7</td>
<td>D</td>
<td>7 days -36 months</td>
<td>Temperature (rectal) ≥38.0°C and no localising signs of infection in history or at physical examination</td>
<td>Bacteraemia, pyelonephritis, lobar pneumonia, meningitis, osteoarthritis</td>
</tr>
<tr>
<td>Garra 2003</td>
<td>Prospective, cross sectional</td>
<td>Emergency department, USA</td>
<td>181</td>
<td>21.6</td>
<td>C</td>
<td>29-56 days</td>
<td>Temperature (rectal) ≥38.1°C</td>
<td>Urinary tract infection, bacteraemia, bacterial meningitis, pneumonia, bacterial enteritis</td>
</tr>
<tr>
<td>Hsiao 2006</td>
<td>Prospective, cross sectional</td>
<td>Emergency department, USA</td>
<td>429</td>
<td>10.3</td>
<td>C</td>
<td>57-180 days</td>
<td>Temperature (rectal) ≥37.9°C</td>
<td>Bacteraemia, urinary tract infection</td>
</tr>
<tr>
<td>Nademi 2001</td>
<td>Prospective, cross sectional, consecutive</td>
<td>Paediatric assessment unit, UK</td>
<td>141</td>
<td>29.1</td>
<td>D</td>
<td>0-16 years</td>
<td>Temperature ≥38°C</td>
<td>Sepsis, meningitis, toxic shock syndrome, brain abscess, pneumonia, urinary tract infection, ischiorectal abscess, appendicitis</td>
</tr>
<tr>
<td>Thayyil 2005</td>
<td>Prospective, cross sectional, consecutive</td>
<td>Paediatric department, UK</td>
<td>72</td>
<td>11.1</td>
<td>D</td>
<td>1-36 months</td>
<td>Temperature ≥39°C without localising signs</td>
<td>Bacteraemia, bacterial meningitis, acute pyelonephritis</td>
</tr>
<tr>
<td>Trautner 2006</td>
<td>Prospective, cross sectional</td>
<td>Emergency department, USA</td>
<td>103</td>
<td>19.4</td>
<td>C</td>
<td>&lt;17 years</td>
<td>Temperature (rectal) ≥41.1°C</td>
<td>Bacterial meningitis, sepsis, bacteraemia, urinary tract infection, bacterial gastroenteritis, pneumonia</td>
</tr>
</tbody>
</table>

Confidence intervals reported are calculated on the basis of the standard error of a proportion using Stata v9.2. In case of an empty cell in the 2×2 table, we added 0.5 to the cell to compute likelihood ratios. All studies were categorised according to setting by using prevalence as a proxy: less than 5%, low prevalence settings; 5-20%, intermediate; and more than 20%, high. We report both the pre-test and post-test probabilities of serious infection for each study, choosing the cut-off levels that were most commonly reported.
across studies, and we report the results from each study once only. We show the change in probability of disease after applying the test. The probability of disease before testing equals the prevalence of serious infections in that study, the probability of disease after a positive test result equals the positive predictive value, and the probability of disease after a negative test result equals 1−negative predictive value.

Meta-analysis was carried out when at least four different studies were available for a particular laboratory test, using the bivariate random effects method in Stata v9.2, which measures and accounts for any statistical heterogeneity in sensitivity and specificity between studies on the logit scale. We used a minimum of four studies, for a reliable estimate of the correlation between sensitivity and specificity. No imputation of empty cells was applied here as this is accounted for in the method. We did not include overlapping data in the meta-analysis, by selecting only one cut-off value for each study. In addition, since some of the results reported by Galetto-Lacour 2001 and 2003 were included in Galetto-Lacour 2008, we selected the 2008 publication when available and used the other studies only for those tests that were not reported in the 2008 paper.

For C reactive, procalcitonin, and white blood cell count, some studies reported sensitivity and specificity at multiple thresholds. We pooled these data using the bivariate method of Dukic and Gatsonis, which is an extension of the usual bivariate method but allows the use of multiple threshold values from each study. The method was implemented in R, with point wise confidence intervals for the sensitivity at a given threshold calculated by simulation.

RESULTS

The electronic search identified 1860 articles (fig 1). An additional 79 articles were identified in the snowballing strategy. In total, 255 articles were selected for review in full text, of which 36 met the selection criteria. Twenty one of these studies focused on clinical features only or were restricted to specific disease outcomes (for example, meningitis) and so were not included in the analysis reported here. In addition, one study was excluded because the 2×2 tables could not be constructed, despite the authors being contacted. Table 1 shows the full details of the 14 articles reporting on 13 different studies.

The quality of the 14 articles was modest (see web extra annex 2): none received an A rating. One was rated B, six were rated C, and seven were rated D. Only one study explicitly mentioned blind reading of the reference standard; this item was scored as unclear in eight studies. Only three studies reported indeterminate or intermediate results.

Most of the studies were carried out in emergency departments, with one study done in a paediatric department and one in a paediatric assessment unit. All studies selected participants on the presence of fever. The median prevalence of serious infection was 20.5% (range 4.5-29.3%). The age of eligible children also varied among the studies, with four studies including infants only (<3 months) and two studies including older participants, up to 16 years of age. The reported outcomes included bacteraemia (14 studies), sepsis (four), meningitis (13, bacterial only in eight cases), pneumonia (11), and urinary tract infection (14), and in some cases additional infections such as bone or joint infections (eight), bacterial infections in that study. The probability of disease after a positive test result equals the positive predictive value, and the probability of disease after a negative test result equals 1−negative predictive value.

Meta-analysis was carried out when at least four different studies were available for a particular laboratory test, using the bivariate random effects method in Stata v9.2, which measures and accounts for any statistical heterogeneity in sensitivity and specificity between studies on the logit scale. We used a minimum of four studies, for a reliable estimate of the correlation between sensitivity and specificity. No imputation of empty cells was applied here as this is accounted for in the method. We did not include overlapping data in the meta-analysis, by selecting only one cut-off value for each study. In addition, since some of the results reported by Galetto-Lacour 2001 and 2003 were included in Galetto-Lacour 2008, we selected the 2008 publication when available and used the other studies only for those tests that were not reported in the 2008 paper.

For C reactive, procalcitonin, and white blood cell count, some studies reported sensitivity and specificity at multiple thresholds. We pooled these data using the bivariate method of Dukic and Gatsonis, which is an extension of the usual bivariate method but allows the use of multiple threshold values from each study. The method was implemented in R, with point wise confidence intervals for the sensitivity at a given threshold calculated by simulation.

RESULTS

The electronic search identified 1860 articles (fig 1). An additional 79 articles were identified in the snowballing strategy. In total, 255 articles were selected for review in full text, of which 36 met the selection criteria. Twenty one of these studies focused on clinical features only or were restricted to specific disease outcomes (for example, meningitis) and so were not included in the analysis reported here. In addition, one study was excluded because the 2×2 tables could not be constructed, despite the authors being contacted. Table 1 shows the full details of the 14 articles reporting on 13 different studies.

The quality of the 14 articles was modest (see web extra annex 2): none received an A rating. One was rated B, six were rated C, and seven were rated D. Only one study explicitly mentioned blind reading of the reference standard; this item was scored as unclear in eight studies. Only three studies reported indeterminate or intermediate results.

Most of the studies were carried out in emergency departments, with one study done in a paediatric department and one in a paediatric assessment unit. All studies selected participants on the presence of fever. The median prevalence of serious infection was 20.5% (range 4.5-29.3%). The age of eligible children also varied among the studies, with four studies including infants only (<3 months) and two studies including older participants, up to 16 years of age. The reported outcomes included bacteraemia (14 studies), sepsis (four), meningitis (13, bacterial only in eight cases), pneumonia (11), and urinary tract infection (14), and in some cases additional infections such as bone or joint infections (eight), bacterial infections in that study. The probability of disease after a positive test result equals the positive predictive value, and the probability of disease after a negative test result equals 1−negative predictive value.

Meta-analysis was carried out when at least four different studies were available for a particular laboratory test, using the bivariate random effects method in Stata v9.2, which measures and accounts for any statistical heterogeneity in sensitivity and specificity between studies on the logit scale. We used a minimum of four studies, for a reliable estimate of the correlation between sensitivity and specificity. No imputation of empty cells was applied here as this is accounted for in the method. We did not include overlapping data in the meta-analysis, by selecting only one cut-off value for each study. In addition, since some of the results reported by Galetto-Lacour 2001 and 2003 were included in Galetto-Lacour 2008, we selected the 2008 publication when available and used the other studies only for those tests that were not reported in the 2008 paper.

For C reactive, procalcitonin, and white blood cell count, some studies reported sensitivity and specificity at multiple thresholds. We pooled these data using the bivariate method of Dukic and Gatsonis, which is an extension of the usual bivariate method but allows the use of multiple threshold values from each study. The method was implemented in R, with point wise confidence intervals for the sensitivity at a given threshold calculated by simulation.
The laboratory tests measured C reactive protein (five studies), procalcitonin (three), erythrocyte sedimentation rate (one), interleukins (two), white blood cell count (seven), absolute neutrophil count (two), band count (three), and left shift (one).

### Inflammatory markers

Figure 2 shows that all the inflammatory markers assessed offer important diagnostic information, although the interleukin tests performed less well than the C reactive protein and procalcitonin tests. The positive likelihood ratios for the procalcitonin test ranged from 1.73 to 3.11 and the negative likelihood ratios ranged from 0.08 to 0.35. The positive likelihood ratios for the C reactive protein test ranged from 2.40 to 3.79 and the negative likelihood ratios ranged from 0.25 to 0.61. The positive likelihood ratios for the interleukin tests ranged from 1.89 to 2.74 and the negative likelihood ratios ranged from 0.33 to 0.77.

Six studies included more selected patient populations—that is, children with fever without source, either referred14–17 or not referred.14–21 All three of the studies that evaluated procalcitonin concerned children with fever without a source, whereas for C reactive protein this population was used in three of five studies and for white blood cell count in three of seven studies. The limited number of studies did not allow formal statistical testing for differences in diagnostic value between these groups.

### White blood cell counts

Figure 3 shows that white blood cell count assays provide less diagnostic value than tests for either C reactive protein or procalcitonin. They provide minimal diagnostic value in ruling out serious infection: the minimum negative likelihood ratio was 0.61, with wide 95% confidence intervals in most studies crossing unity. They provided some diagnostic value for ruling in serious infection, with positive likelihood ratios ranging from 0.87 to 2.43. Although some studies reported imprecise results with wide confidence intervals, two studies included over 400 children and therefore provide the most precise estimates: Hsiao et al22 reported a positive likelihood ratio of 2.43 (95% confidence interval 1.73 to 3.43) and negative likelihood ratio of 0.61 (0.44 to 0.83) for white blood cell counts of 15×10⁹/L or more, and Andreola et al16 reported a positive likelihood ratio of 2.08 (1.58 to 2.75) and a negative likelihood ratio of 0.65 (0.52 to 0.80) for white blood cell counts greater than 15×10⁹/L. Strikingly, one study reported paradoxical results for white blood cell counts: the likelihood of serious infection was lower in children with a white blood cell count of 15×10⁹/L or more, and Andreola et al16 reported a positive likelihood ratio of 2.08 (1.58 to 2.75) and a negative likelihood ratio of 0.65 (0.52 to 0.80) for white blood cell counts greater than 15×10⁹/L. Strikingly, one study reported paradoxical results for white blood cell counts: the likelihood of serious infection was lower in children with a white blood cell count of 15×10⁹/L or more.21 The band count seems the best performing white blood cell marker, with two studies reporting a positive likelihood ratio of around 3, but the rule out value was still poor (with the 95% confidence intervals for a negative likelihood ratio in two of the three studies including unity). One study showed that white blood cell count did not offer additional...
the clinical prediction rule, a goodness of fit was reported (Hosmer-Lemeshow test for the laboratory model $P=0.3$), the area under the curve for the clinical and laboratory model was 0.86 (95% confidence interval 0.82 to 0.90). This rule was not validated in an independent dataset (the clinical part of the prediction rule was derived in the merged derivation and validation sets).

The only clinical feature explicitly considered in these two best performing prediction rules was fever, which was an inclusion criterion for all the studies reported here. Neither study reported validation or other metrics for performance of the prediction rule.

The prediction rule reported by Bleeker et al. gives an estimate of the added value that blood tests provide for patients testing positive or negative to a series of clinical features. This suggests that blood testing provides greater ability to rule out serious infection in children at higher risk based on clinical features. In these children, results of a combination of white blood cell count, C reactive protein, and urinalysis will lower the probability from 42% to 15% when the test result is negative, but the probability is not increased substantially when the test result is positive (54%). On the other hand, in patients testing negative on the clinical prediction rule and therefore at lower risk of serious infection, the blood tests moderately lower the probability (from 12% to 4%) if the results are negative and moderately increase the probability if positive (to 31%). For this clinical prediction rule, a goodness of fit was reported (Hosmer-Lemeshow test for the laboratory model $P=0.3$), the area under the curve for the clinical and laboratory model was 0.86 (95% confidence interval 0.82 to 0.90). This rule was not validated in an independent dataset (the clinical part of the prediction rule was derived in the merged derivation and validation sets).

For completeness figure 5 also reports four clinical prediction rules that include more invasive investigations as well as blood tests and clinical assessment (cerebrospinal fluid variables and chest radiography). These prediction rules, such as the Philadelphia protocol and Milwaukee protocol, are usually applied in infants aged less than 3 months presenting to the emergency department with fever. The plots show they provide little diagnostic value in ruling in serious infection (positive likelihood ratio ranging from 1.27 to 1.70 but...
are helpful at ruling it out (negative likelihood ratio ranging from 0.04 to 0.15).

Meta-analyses

Standard bivariate random effects meta-analysis was possible for C reactive protein tests and the invasive protocols in infants (table 2). Meta-analysis for C reactive protein yielded a pooled positive likelihood ratio of 3.15 (2.67 to 3.71) and a pooled negative likelihood ratio of 0.33 (0.22 to 0.49) across all cut-offs values (five studies, 1379 children). For the invasive protocols the pooled positive likelihood ratio was 1.40 (1.27 to 1.55) and the pooled negative likelihood ratio was 0.06 (0.02 to 0.19).

Meta-analyses generated by the Dukic and Gatsonis method were possible for both C reactive protein and procalcitonin. Meta-analysis of white blood cell count was not possible owing to heterogeneity. Figure 6 shows the results of these analyses. These hierarchical summary receiver operating characteristic curves show the sensitivity and 1−specificity for each possible level of C reactive protein and procalcitonin. The more the curve approaches the left upper corner, the better the diagnostic accuracy, whereas those closer to the diagonal have poorer diagnostic accuracy.

The figure confirms that C reactive protein and procalcitonin have comparable diagnostic accuracy, as the shape of the curves is roughly similar and the confidence intervals are largely overlapping. In addition, most data points for white blood cell count lie close to the centre of the plot, indicating limited diagnostic value—for example, a sensitivity of 50% and a specificity of 60%.

Cut-off levels

Data from the receiver operating characteristic figure were used to select optimal cut-off values to label the test result for C reactive protein and procalcitonin as abnormal. Although heterogeneity prevented reliable estimates of pooled likelihood ratios for each cut-off value, the shapes of the curves in the figure illustrate that there is a trade-off in sensitivity and specificity according to the cut-off level selected. To rule in serious infection in a feverish child, cut-off levels of 80 mg/L for C reactive protein or 2 ng/mL for procalcitonin both provide specificity of more than 90% but a sensitivity of 40-50%, which in the original studies corresponded to a positive likelihood ratio of 8.4 (C reactive protein) and 3.6 to 13.7 (procalcitonin), and negative likelihood ratio of 0.57 (C reactive protein) and 0.54 to 0.58 (procalcitonin). To rule out serious infection effectively, cut-off levels of 20 mg/L for C reactive protein or 0.5 ng/mL for procalcitonin may be a better choice, with sensitivity more than 80% but specificity 70% (corresponding to a negative likelihood ratio of 0.19 to 0.25 and 0.08 to 0.25, respectively, in the original studies).

Three studies compared procalcitonin with C reactive protein directly.14 16 21 One study found no statistically significant difference in the areas under the curve for C reactive protein and procalcitonin (P=0.75).
Invasive protocols

<table>
<thead>
<tr>
<th>Test</th>
<th>No of studies/No of children</th>
<th>% sensitivity (95% CI)</th>
<th>% specificity (95% CI)</th>
<th>Likelihood ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C reactive protein</td>
<td>5/1379</td>
<td>75.1 (62.7 to 84.5)</td>
<td>76.1 (71.0 to 80.6)</td>
<td>3.15 (2.67 to 3.71)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.33 (0.22 to 0.49)</td>
</tr>
<tr>
<td>Invasive protocols</td>
<td>4/1884</td>
<td>98.2 (94.5 to 99.4)</td>
<td>99.9 (96.9 to 99.9)</td>
<td>1.40 (1.27 to 1.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.06 (0.02 to 0.19)</td>
</tr>
</tbody>
</table>

DISCUSSION

In this systematic review few studies assessed the usefulness of blood tests in identifying serious infection in children presenting to an emergency setting, despite it being an endemic clinical activity in all health systems. The studies we did find were of modest methodological quality and none had been carried out in a low prevalence primary care setting. Moreover, it was difficult to extract the diagnostic added value of testing—that is, in what clinical situations testing is likely to contribute substantially to diagnostic decision making—on the basis of the data reported. Few studies had made explicit the asymmetry of diagnostic value in ruling in and ruling out serious infections, and most reported findings in terms of sensitivity and specificity, which provide little useful information to clinicians interested in the diagnostic value of tests in making clinical decisions.

However, the data suggest that: the tests providing most diagnostic value in feverish children are C reactive protein and procalcitonin, with no clear evidence from our data that one is better than the other; the optimal cut-off point for C reactive protein and procalcitonin depends on whether the main clinical focus is ruling in or ruling out serious infection, and the best performing clinical prediction rule uses the actual level to derive a predictive score; and white blood cell indicators probably provide some diagnostic value in ruling in serious infection, but less than the inflammatory markers, and have no value at ruling it out.

Although tests for C reactive protein and procalcitonin have similar diagnostic properties they might be of different value at different points in the course of illness, as theoretically procalcitonin levels increase earlier than those for C reactive protein. This is supported by the inclusion of both C reactive protein and procalcitonin in the best performing clinical prediction rule, although it is unclear whether doing both tests is a cost effective approach.

The results of an earlier review, which was limited to C reactive protein in children with fever, used slightly different methods but provided similar results: pooled positive likelihood ratio 3.64 (95% confidence interval 2.99 to 4.43) and negative likelihood ratio 0.29 (0.22 to 0.40) compared with a positive likelihood ratio of 3.15 (2.67 to 3.71) and a negative likelihood ratio of 0.33 (0.22 to 0.49) in our study. Other reviews are less comparable because they included both adults and children or used different outcomes.

Robustness of the findings

The review methodology was robust, involving an extensive search, and we should have identified all existing good quality studies, which in aggregate included 3981 children.

The reliability of the data is limited by the methodological quality of the studies included in the review, but the modest QUADAS scores reported in part reflect the limitations of undertaking pragmatic studies in a clinical setting. The ideal diagnostic accuracy study would include all eligible patients consecutively, carry out the same index test and reference standard in all patients identically, and read both tests blinded from each other. It is difficult to meet the highest methodological standard for confirming the final diagnosis in a study in routine practice where it is not possible for all children included to have a full investigation. Even so, the quality of the reporting of the methodology of many of the studies was not optimal, although several C rated studies met the criterion of as good as can be done in routine practice. The importance of large scale pragmatic studies should not be down rated by setting quality criteria that are not feasible. However, the fact that all the studies identified were done on populations of febrile children in settings where serious infection was relatively common (typically 1 in 5 children had serious infection) does limit generalisation to primary care settings.

care settings where typically 1 in 150 children have a serious infection.26

Comparison with developing countries
Our review excluded studies from developing countries because of differences in prevalence; microbiological causes; comorbidities, such as HIV and malnutrition; and more advanced stage of disease at presentation. For these reasons, the diagnostic characteristics of laboratory tests in these settings are likely to be different and should be treated separately.

In general, focus is on clinical signs and symptoms. The World Health Organization has issued the integrated management of childhood illness guidelines to improve recognition and treatment of the five most important illnesses in children: pneumonia, diarrhoea, malaria, measles, and malnutrition. The guidelines emphasise triage based on clinical signs and symptoms, and consequent treatment at a local facility or referral to a regional facility. The guidelines currently do not include specific recommendations on laboratory tests.

Strategic implications
The evidence base for investigations in children needs strengthening. One of the problems is that most of the studies done to date do not show the way the investigation is carried out in practice and often provide little evidence of the added value of different (and serially applied) diagnostic strategies at each stage in the decision making process. Our previous review of the diagnostic value of presenting symptoms and signs showed a substantial diagnostic gap between the predictive value achievable by clinical features without further investigation and the threshold of risk of serious infection that most hospital clinicians would wish to apply to justify hospital admission.1 This gap is currently filled by the application of poorly defined clinical “gut feeling” and similarly poorly defined strategies for diagnostic safety netting,27 so perhaps it is not surprising that missed cases of serious infection through failed investigations remain common.28 In addition, clinicians may have a low threshold for immediate prescription of antibiotics, especially in young children, to avoid missing a serious infection.29-30

Providing robust evidence to help parents and clinicians differentiate children with serious infections from those with minor illnesses is important for children’s health services. It is paradoxical that acute infections continue to generate such large numbers of patient consultations (in primary and secondary care), yet the incidence of serious infections has declined, partly as a result of vaccination.31 With increased pressures on health service resources, more robust evidence is needed to identify those few children who need referral or admission. At present up to one third of short stay admissions in infants are for minor illnesses.32

The evidence presented here suggests that the role of inflammatory markers is a key issue, but the potential symbiotic value of recording vital signs must also be assessed.33 As such, studies in primary care are needed to determine the diagnostic value of point of care testing for C reactive protein and procalcitonin in combination with clinical signs and vital signs (as well as their feasibility, acceptability, and cost effectiveness). Such studies are, however, difficult to design because of the large sample sizes involving many practices and doctors, and consequent large budgets. In addition, a diagnostic trial could evaluate different diagnostic strategies in primary and secondary care, with risk stratification based on clinical features and vital signs, and laboratory testing (preferably point of care) in selected patients. The trial could evaluate outcomes such as number of referrals to secondary care, further investigations in secondary care, admission to hospital, missed or delayed diagnosis of a serious infection, prescription of antibiotics, as well as cost effectiveness of different strategies.

Immediate clinical implications
This review provides evidence that measuring the white blood cell count provides limited diagnostic help in identifying serious infection in children. In a primary care setting, where taking blood is invasive and likely to be of high marginal cost, no case exists for its use. The inflammatory markers have more diagnostic value but the low likelihood of serious infection in primary care means that whatever cut-off value is applied the test result is unlikely to change the probability of serious infection to the extent that it would impact on a clinical decision to treat, further investigate, or admit to hospital. Moreover, this review did not identify any studies that directly assessed the diagnostic value of either white blood cell count or inflammatory markers in identifying serious childhood infection in primary care.

The situation is somewhat different in a hospital emergency department or a general practice out of hours assessment centre, as the children presenting are more likely to have serious infection or other serious illness. The higher pre-test probability of serious disease means that the test results will have a greater absolute effect on post-test probabilities. However, measurement of C reactive protein and procalcitonin will better inform clinical decision making than the white blood cell count in diagnosing serious infection, as long as appropriate thresholds are applied. Combining blood tests—that is, C reactive protein and white blood cell count—seems not very helpful. At present the test for C reactive protein is cheaper and easier to carry out at point of care than the test for procalcitonin, thus making it the preferred choice. In addition, point of care tests for C reactive protein have been shown to generally correlate well with classic laboratory tests in studies in primary care and emergency departments.34-36

With the exception of the invasive protocols, such as the Milwaukee protocol, that are done routinely in feverish infants and that show great value in excluding serious infections, most clinicians will assess children first and selectively request laboratory tests depending...
White blood cell counts are less useful. C reactive protein and procalcitonin may be useful measures, but different cut-off values should be used for ruling in or ruling out serious infections. White blood cell counts are less useful. No evidence from primary care was identified.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi-disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Copyright: This work was funded by the Health Technology Assessment and is therefore subject to standard arrangements for use of Crown copyright material by other organisations.

Data sharing: No additional data available.


6 Van der Meer V, Neven AK, van den Broek PI, Assendelft WJ. Diagnostic value of C-reactive protein in infections of the lower respiratory tract: systematic review. BMJ 2005;331:26.


23 Trautner BW, Cavness AC, Gerlacher GR, Demmler G, Macias CG. Prospective evaluation of the risk of serious bacterial infection in children who present to the emergency department with hyperpyrexia (temperature of 106 degrees F or higher). *Pediatrics* 2006;118:34-40.

Accepted: 6 April 2011