Accuracy of procalcitonin for sepsis diagnosis in critically ill patients: systematic review and meta-analysis

Benjamin M P Tang, Guy D Eslick, Jonathan C Craig, Anthony S McLean

Procalcitonin is widely reported as a useful biochemical marker to differentiate sepsis from other non-infectious causes of systemic inflammatory response syndrome. In this systematic review, we estimated the diagnostic accuracy of procalcitonin in sepsis diagnosis in critically ill patients. 18 studies were included in the review. Overall, the diagnostic performance of procalcitonin was low, with mean values of both sensitivity and specificity being 71% (95% CI 67–76) and an area under the summary receiver operator characteristic curve of 0.78 (95% CI 0.73–0.83). Studies were grouped into phase 2 studies (n=14) and phase 3 studies (n=4) by use of Sackett and Haynes' classification. Phase 2 studies had a low pooled diagnostic odds ratio of 7.79 (95% CI 5.86–10.35). Phase 3 studies showed significant heterogeneity because of variability in sample size (meta-regression coefficient −0.592, p=0.017), with diagnostic performance upwardly biased in smaller studies, but moving towards a null effect in larger studies. Procalcitonin cannot reliably differentiate sepsis from other non-infectious causes of systemic inflammatory response syndrome in critically ill adult patients. The findings from this study do not lend support to the widespread use of the procalcitonin test in critical care settings.

Introduction

Sepsis is the leading cause of mortality in critically ill patients.1 Delay in diagnosis and treatment often results in rapid progression to circulatory collapse, multiple organ failure, and eventually death.2,3 Therefore, accurate and timely diagnosis will limit morbidity, reduce costs, and improve patients’ outcome.4,5

The diagnosis of sepsis is difficult, because clinical signs of sepsis often overlap with other non-infectious causes of systemic inflammation.6–8 These signs include tachycardia, leukocytosis, tachypnoea, and pyrexia, which are collectively termed a systemic inflammatory response syndrome (SIRS). SIRS is very common in critically ill patients, being found in various conditions including trauma, surgery, and hypoxic injuries.6–11 Microbiological culture can be used to distinguish sepsis from non-infectious conditions. However, this method lacks sensitivity and specificity, and there is often a substantial time delay.12

Procalcitonin, a 116-aminoacid peptide involved as a precursor in calcium homeostasis, has been studied as a marker to differentiate sepsis from other non-infectious causes of SIRS. Early studies were encouraging,13–16 and procalcitonin has been proposed as a diagnostic marker to be included in the international definition of sepsis.17 However, more recent studies have produced conflicting results.18–24 Furthermore, many studies included patients who did not have SIRS or who were not critically ill. This has added further uncertainty in assessing the diagnostic accuracy of procalcitonin in the critical care setting. The aim of this review was therefore to systematically and quantitatively evaluate all the published studies that assessed the diagnostic use of procalcitonin in critical care settings.

Methods

Data source

We searched Medline, Embase, and Current Contents from January, 1966, to November, 2005, for all studies of diagnostic accuracy of procalcitonin for sepsis. The search strategy used medical subject heading terms and text words, including the following: “procalcitonin”; “sepsis”, “sepsis syndrome”, “septicemia”, “infection”, “systemic inflammatory response syndrome”, and “SIRS”; and “sensitivity”, “specificity”, “predictive value”, “likelihood ratio”, “review”, “meta-analysis”, “false positive”, and “false negative”.

The reference lists of each primary study were searched for additional publications. Further searches were done by manually reviewing abstract booklets, conference proceedings, and review articles. Investigators were contacted for further study details if needed. No language restriction was used and all foreign language publications were translated.

Study eligibility

We included all studies that met the following criteria: assessed the diagnostic accuracy of procalcitonin for sepsis; provided sufficient information to construct the 2x2 contingency table; and had a well-defined reference standard for the target condition (sepsis), which included use of accepted definitions by the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference,1 and confirmed the presence of infection by microbiological culture.

Studies were excluded if they included patients who did not have SIRS or were not critically ill; included too narrow spectrum of patients, such as abdominal sepsis or septic shock; were duplicated studies; were paediatric studies; were limited to very restrictive subgroups, such as cardiac surgery, pancreatitis, meningitis, or burns; or were risk stratification or prognosis studies.

Data extraction

Two reviewers (BMPT, GDE) independently abstracted data in each study to obtain information on year of publication, country of origin, clinical setting, patients’
and colleagues,\textsuperscript{26} and the QUADAS tool.\textsuperscript{27,28} Details of the methodological assessment are shown in table 1.

The diagnostic odds ratio (OR) expresses the magnitude by which the odds of a positive result in a patient with sepsis is relative to the odds in a patient without sepsis: $\text{OR} = \frac{\text{sensitivity}}{(1-\text{sensitivity})}/\frac{(1-\text{specificity})}{\text{specificity}}$. The diagnostic OR is a measure of overall accuracy and has the advantage of being independent of prevalence. Diagnostic OR is the ratio of the odds of a positive result in a patient with sepsis compared with a patient without sepsis: $\text{OR} = \frac{[\text{sensitivity}/(1-\text{sensitivity})]/(1-\text{specificity})/\text{specificity}}{[\text{specificity}]/(1-\text{specificity})}$. The diagnostic OR is a measure of overall accuracy and has the advantage of being independent of prevalence. The diagnostic OR is the ratio of the odds of a positive result in a patient with sepsis compared with a patient without sepsis: $\text{OR} = \frac{\text{sensitivity}}{(1-\text{sensitivity})}/\frac{(1-\text{specificity})}{\text{specificity}}$. The diagnostic OR is a measure of overall accuracy and has the advantage of allowing the inclusion of covariates to examine heterogeneity in a regression model.\textsuperscript{30} Pooling of the summary indices was done using DerSimonian and Laird’s random-effects model.\textsuperscript{31} Each study was weighted by use of an inverse variance method.

To detect heterogeneity, the likelihood ratios and diagnostic ORs were graphically displayed using forest plots and analysed using Cochran’s $Q$ test. A $p$ value of less than 0.05 by Cochran’s $Q$ test indicated significant heterogeneity. To quantify the extent of heterogeneity, the $I^2$ statistic was used to measure the percentage of variability among summary indices that were caused by heterogeneity rather than chance. A study with an $I^2$ greater than 50% indicated substantial heterogeneity.

We constructed summary receiver operator characteristic (SROC) curves to summarise the study results, by use of a regression model described by Littenberg and Moses.\textsuperscript{32} In this method, the true-positive and false-positive rates of each study were logarithmically transformed and calculated in a regression model. The data were then back-transformed into the SROC space. A smoothed curve was then fitted across studies to represent the relation between sensitivity and the proportion of false positives (1–specificity).

To ensure that variation in the diagnostic threshold did not affect the shape of the SROC curve, the threshold effect was tested using the regression equation $D=a+bs$, where $D$ is the log of the diagnostic OR and $S$ is a measure of the diagnostic threshold. Estimation of the variables $a$ and $b$ was then done using a least-squares method, weighted by inverse variance. The absence of a threshold effect was indicated by $b=0$.

A $Q^*$ point on the SROC curve was used to obtain the maximum joint sensitivity and specificity. The $Q^*$ point is the intersection between a symmetrical SROC curve and the antidiagonal line, at which sensitivity equals specificity. This point represents a single-number summary of the

<table>
<thead>
<tr>
<th>Methodological variable</th>
<th>Information required in each study</th>
<th>Studies that met criteria (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did investigators use additional information (other than consensus definition of sepsis and microbiological culture) to confirm diagnosis, thus minimising misclassification bias?</td>
<td>Using all available information to diagnose sepsis/SIRS, including imaging studies, response to antibiotics, necropsy reports, and surgical findings</td>
<td>14</td>
</tr>
<tr>
<td>Was there a time delay between the index test and reference test (disease progression bias)?</td>
<td>Both procalcitonin and reference test to be done at the same time</td>
<td>18</td>
</tr>
<tr>
<td>Did the result of index test influence whether patients receive reference test (work-up bias)?</td>
<td>All patients should receive reference test regardless of procalcitonin test results</td>
<td>18</td>
</tr>
<tr>
<td>Were different reference tests used in patients (differential verification bias)?</td>
<td>Consistent use of international consensus criteria to diagnose sepsis in all patients</td>
<td>18</td>
</tr>
<tr>
<td>Was the interpretation of the reference test made without the knowledge of the index test (blinding)?</td>
<td>Diagnosis of sepsis/SIRS was made independent of the result of procalcitonin test</td>
<td>8</td>
</tr>
<tr>
<td>Description of reference test</td>
<td>Sufficient details provided in how the diagnosis was made</td>
<td>18</td>
</tr>
<tr>
<td>Description of index test</td>
<td>Sufficient details provided in how the procalcitonin was measured</td>
<td>18</td>
</tr>
<tr>
<td>Description of study population</td>
<td>Sufficient details provided for the case mix and demographic information of the patients enrolled</td>
<td>18</td>
</tr>
<tr>
<td>Method of recruitment</td>
<td>Patients were prospectively or consecutively recruited</td>
<td>18</td>
</tr>
</tbody>
</table>

SIRS=systemic inflammatory response syndrome.

Table 1: Quality assessment of the 18 studies included, by methodological variable.

demographics, sample size, diagnostic cut-off points, and disease prevalence. Each reviewer extracted the data to construct a 2×2 table. Any disagreements were resolved by consensus.

Quality assessment
The methodological quality of each study was assessed by a checklist, by use of adapted criteria from the Cochrane Collaboration guidelines,\textsuperscript{25} a study by Lijmer and colleagues,\textsuperscript{29} and the QUADAS tool.\textsuperscript{27,28} Details of the methodological assessment are shown in table 1.

Statistical analysis
Studies were grouped according to Sackett and Haynes’ classification of diagnostic studies.\textsuperscript{30} In this classification, phase 1 studies are those that compare the difference in test results between patients with the target disorder and healthy individuals. Phase 2 studies are those that examine how the index test discriminates between patients with and without the target disorder. Phase 3 studies are those that assess the test’s real-life performance in patients suspected to have the disorder. For each study, positive and negative likelihood ratios and a diagnostic odds ratio (OR) were calculated. The likelihood ratio expresses the magnitude by which the probability of sepsis in a given patient is modified by the results of the procalcitonin test. It incorporates both sensitivity and specificity and has the advantage of being independent of prevalence. The diagnostic OR is the ratio of the odds of a positive result in a patient with sepsis compared with a patient without sepsis: $\text{OR} = \frac{[\text{sensitivity}/(1-\text{sensitivity})]/(1-\text{specificity})/\text{specificity}}{[\text{specificity}]/(1-\text{specificity})}$. The diagnostic OR is a measure of overall accuracy and has the advantage of allowing the inclusion of covariates to examine heterogeneity in a regression model.\textsuperscript{30} Pooling of the summary indices was done using DerSimonian and Laird’s random-effects model.\textsuperscript{31} Each study was weighted by use of an inverse variance method.

To detect heterogeneity, the likelihood ratios and diagnostic ORs were graphically displayed using forest plots and analysed using Cochran’s $Q$ test. A $p$ value of less than 0.05 by Cochran’s $Q$ test indicated significant heterogeneity. To quantify the extent of heterogeneity, the $I^2$ statistic was used to measure the percentage of variability among summary indices that were caused by heterogeneity rather than chance. A study with an $I^2$ greater than 50% indicated substantial heterogeneity.

We constructed summary receiver operator characteristic (SROC) curves to summarise the study results, by use of a regression model described by Littenberg and Moses.\textsuperscript{32} In this method, the true-positive and false-positive rates of each study were logarithmically transformed and calculated in a regression model. The data were then back-transformed into the SROC space. A smoothed curve was then fitted across studies to represent the relation between sensitivity and the proportion of false positives (1–specificity).

To ensure that variation in the diagnostic threshold did not affect the shape of the SROC curve, the threshold effect was tested using the regression equation $D=a+bs$, where $D$ is the log of the diagnostic OR and $S$ is a measure of the diagnostic threshold. Estimation of the variables $a$ and $b$ was then done using a least-squares method, weighted by inverse variance. The absence of a threshold effect was indicated by $b=0$.

A $Q^*$ point on the SROC curve was used to obtain the maximum joint sensitivity and specificity. The $Q^*$ point is the intersection between a symmetrical SROC curve and the antidiagonal line, at which sensitivity equals specificity. This point represents a single-number summary of the

http://infection.thelancet.com Vol 7 March 2007
test performance and has the advantage of being less affected than other parameters by heterogeneity.32,33

To explore sources of heterogeneity among studies, the Littenberg-Moses method32 was extended by adding covariates to the model. The covariates included spectrum characteristics (eg, study setting, prevalence), clinical and demographic variables (eg, disease severity, age), and methodological features (eg, sample size).

Publication bias was examined visually by inspecting funnel plots and statistically by using Egger’s regression model.34 If publication bias was present, the effect of such bias on the final summary estimate was assessed by using the trim and fill method.35 This method imputes the missing studies and re-calculates a new summary estimate. The difference between the calculated and observed value was then used to determine the effect of bias on the diagnostic performance of the test.

**Results**

**Study characteristics**

We retrieved 672 abstracts, of which 39 were considered potentially suitable. After full text review, 21 studies were excluded (figure 1): one had no SIRS patients in the control group,36 four included patients who were not critically ill,37–40 two were case-control studies,41,42 three used a different reference standard,43–45 nine could not generate 2×2 tables,14,46–53 and two had too narrow a spectrum of patients.54,55 In total, 18 studies were included in the final analysis. Studies were grouped according to Sackett and Haynes’ classification29 for diagnostic studies: 14 phase 2 studies (group 1), four phase 3 studies (group 2). Details of all 18 studies are shown in table 2.

**Table 2: Characteristics of the studies included (2097 patients)**

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Setting</th>
<th>Patients (n)</th>
<th>Mean age (years)</th>
<th>Cut-off point (ng/mL)</th>
<th>Study design</th>
<th>Prevalence of sepsis</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 studies (1602 patients)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aikawa et al56</td>
<td>2005</td>
<td>Japan</td>
<td>Emergency department</td>
<td>176</td>
<td>47</td>
<td>0.5 PR</td>
<td>51%</td>
<td>0.64</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Al-Nawas et al16</td>
<td>1996</td>
<td>Germany</td>
<td>Hospital ward/ICU</td>
<td>337</td>
<td>..</td>
<td>0.5 PR</td>
<td>36%</td>
<td>0.60</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Baumgarten et al57</td>
<td>2002</td>
<td>Netherlands</td>
<td>ICU</td>
<td>35</td>
<td>..</td>
<td>3 PR</td>
<td>31%</td>
<td>0.55</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Chan et al46</td>
<td>2004</td>
<td>Taiwan</td>
<td>Emergency department</td>
<td>69</td>
<td>65</td>
<td>0.6 PR+CR</td>
<td>54%</td>
<td>0.71</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>Cheval et al59</td>
<td>2000</td>
<td>France</td>
<td>ICU</td>
<td>60</td>
<td>58</td>
<td>20 PR+CR</td>
<td>53%</td>
<td>0.88</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Du et al60</td>
<td>2003</td>
<td>China</td>
<td>ICU</td>
<td>51</td>
<td>65</td>
<td>16 PR+CR</td>
<td>75%</td>
<td>0.80</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Hausfater et al61</td>
<td>2002</td>
<td>France</td>
<td>Emergency department</td>
<td>195</td>
<td>47</td>
<td>0.2 PR</td>
<td>35%</td>
<td>0.62</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Muller et al62</td>
<td>2000</td>
<td>Switzerland</td>
<td>ICU</td>
<td>101</td>
<td>59</td>
<td>10 CR</td>
<td>58%</td>
<td>0.90</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Mokart et al63</td>
<td>2005</td>
<td>France</td>
<td>ICU</td>
<td>50</td>
<td>56</td>
<td>11 PR</td>
<td>47%</td>
<td>0.81</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Selberg et al64</td>
<td>2000</td>
<td>Germany</td>
<td>ICU</td>
<td>33</td>
<td>47</td>
<td>3.3 PR</td>
<td>67%</td>
<td>0.86</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Suprin et al65</td>
<td>2000</td>
<td>France</td>
<td>ICU</td>
<td>95</td>
<td>57</td>
<td>2.00 PR+CR</td>
<td>76%</td>
<td>0.65</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>Tugrul et al66</td>
<td>2002</td>
<td>Turkey</td>
<td>ICU</td>
<td>85</td>
<td>45</td>
<td>1.31 PR</td>
<td>88%</td>
<td>0.73</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Ugarte et al67</td>
<td>1999</td>
<td>Belgium</td>
<td>ICU</td>
<td>182</td>
<td>63</td>
<td>0.6 CR</td>
<td>58%</td>
<td>0.68</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Wanner et al68</td>
<td>2000</td>
<td>Switzerland</td>
<td>ICU</td>
<td>133</td>
<td>40</td>
<td>1.5 PR</td>
<td>34%</td>
<td>0.76</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Group 2 studies (495 patients)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bossink et al69</td>
<td>1999</td>
<td>Netherlands</td>
<td>Hospital ward/ICU</td>
<td>133</td>
<td>60</td>
<td>0.5 CR</td>
<td>45%</td>
<td>0.65</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Gibot et al46</td>
<td>2004</td>
<td>France</td>
<td>ICU</td>
<td>76</td>
<td>60</td>
<td>0.6 PR+CR</td>
<td>62%</td>
<td>0.83</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>Harbarth et al70</td>
<td>2001</td>
<td>Switzerland</td>
<td>ICU</td>
<td>78</td>
<td>54</td>
<td>11 CR</td>
<td>77%</td>
<td>0.97</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Ruokonen et al71</td>
<td>2002</td>
<td>Switzerland</td>
<td>ICU</td>
<td>208</td>
<td>55</td>
<td>0.8 PR+CR</td>
<td>78%</td>
<td>0.68</td>
<td>0.48</td>
<td></td>
</tr>
</tbody>
</table>

ICU=intensive care unit; PR=prospective recruitment; CR=consecutive recruitment; ..=not available.
2097 patients were included in the analysis, with 1452 from intensive care units, 440 from emergency departments, and 205 from hospital wards. Studies included a wide case mix, including cardiac, pulmonary, neurological, gastrointestinal, renal, trauma, and surgical illnesses. SIRS criteria were fulfilled in 2092 patients. The mean age of patients in the studies was 54 years (range of study means 40–65 years). The prevalence of sepsis across studies ranged from 31% to 88%. All studies used LumiTest PCT, a commercially available immunoluminometric assay (Brahms Diagnostica, Berlin, Germany). Test threshold ranged from 0·2 ng/mL to 20 ng/mL.

Quantitative data synthesis

14 studies were included in group 1 (1602 patients). The pooled summary indices showed that the diagnostic performance of procalcitonin was low, with positive likelihood ratio 3·03 (95% CI 2·51–3·65), negative likelihood ratio 0·43 (95% CI 0·37–0·48), and diagnostic OR 7·79 (95% CI 5·86–10·35; figure 2). There was no evidence of a threshold effect (b=0·451, p=0·66). The SROC curve yielded a maximum joint sensitivity and specificity of 73% (95% CI 69–77), an area under the curve of 0·79, and Q* point of 0·73, consistent with low diagnostic accuracy of procalcitonin.

One study had an unusually high summary estimate and accounted for most of the heterogeneity (52·6%). Heterogeneity diminished significantly after this study was excluded (14·7%), thus allowing statistical pooling of the summary estimates. This study was therefore treated as an outlier and the results were reported with the exclusion of this study. However, subsequent sensitivity analysis showed that the pooled summary estimates did not differ significantly with inclusion of the outlier.

Four studies were included in group 2 (495 patients). These studies were highly heterogeneous (Cochran’s Q=21·57, p<0·001), with an I² value of 86·1%. Statistical pooling was therefore not done for this group.

Finally, all 18 studies were pooled. There was no evidence of a threshold effect (b=−0·21, p=0·40). The SROC curve (figure 3) yielded a maximum joint sensitivity and specificity of 71% (95% CI 67–76), an area under the curve of 0·78, and Q* point of 0·72, indicating that the performance of procalcitonin was low even when all studies were combined.

As expected, when pooling all studies, significant heterogeneity was introduced by the group 2 studies (Cochran’s Q=60·21, p<0·001). The source of heterogeneity was explored by univariate meta-regression analysis. Sample size was significant in group 2 as a source of heterogeneity (p=0·017), but only weakly suggestive in group 1 (p=0·09; table 3). None of the variables, such as clinical settings, disease severity, patient demographics, or prevalence, were statistically significant as a source of variability in either group 1 or 2. Within group 2, smaller studies showed a higher diagnostic performance of procalcitonin (eg, a decrease of 40 patients overestimated the relative diagnostic OR by a factor of 1·82). By contrast, the largest study (208 patients) had a diagnostic OR of 1·94 and a 95% CI that included the null effect of 1·0 (figure 4).
Publication bias was detected using Egger’s regression model (p=0.006). Visual inspection of the funnel plot suggested that missing studies were likely to fall to the left of the summary estimate. These studies were then imputed to calculate a new summary estimate (figure 5). The new diagnostic OR was 5.71 (95% CI 3.62–9.03), which was significantly lower than the observed diagnostic OR of 8.71 (95% CI 5.63–13.47). Therefore, the existing studies could have overestimated the diagnostic performance of procalcitonin.

Discussion

The results of this systematic review and meta-analysis indicate that the procalcitonin test cannot accurately distinguish sepsis from SIRS in critically ill adult patients. The study population in this review included a case mix typically seen in medical, surgical, or general intensive care units, emergency departments, and hospital wards. The findings of this review are therefore applicable to common clinical settings in which critically ill patients are managed.

The studies were grouped according to Sackett and Haynes’ classification, which assessed an index test on a continuum of diagnostic uncertainty. This continuum allows a stepwise, systematic progression in diagnostic evaluation from a training set (group 1), in which the index test was developed in an ideal situation, to a validation set (group 2), in which its performance was tested in a more realistic clinical context. Such classification therefore allows clinicians to make a more informed decision when assessing the generalisability of studies.

Most patients (76%) were included in group 1 studies. The diagnostic OR and likelihood ratios were consistently low across most studies in this group. As a general rule, a diagnostic OR of greater than 100 indicates high accuracy, 25–100 indicates moderate accuracy, and less than 25 indicates an unhelpful test. The pooled diagnostic OR of 7.79 showed that the procalcitonin test was unlikely to be helpful in assisting clinical decision making in this group of patients. With a pretest probability of 40% in adult intensive-care-unit patients, use of the procalcitonin test would only raise the post-test probability to 66%. This is insufficient to influence treatment decision (eg, to start antibiotics). Conversely, with a negative likelihood ratio of 0.43, the application of a procalcitonin test would reduce the post-test probability to only 0.23, which is not quite enough to rule out an infection.

The remaining patients (24%) were included in group 2 studies. These studies were the most informative for clinical practice, as they were designed to resemble real-life situations by restricting to patients who were most likely to be encountered by clinicians. Group 2 summary estimates showed lower accuracy and more variability. Sample size gave rise to most of the variability, with smaller studies showing higher summary estimates. Other variables, such as patient age or clinical setting, were likely to have caused variation in the diagnostic performance of procalcitonin. However, the small number of studies (n=4) means that there is a lack of power in detecting these effects. Overall, these data suggest that smaller studies tend to overestimate the effect size, a finding that has been recognised in the diagnostic study literature. A well-designed prospective
study with a larger sample size will be required to address this issue.

The diagnostic accuracy of procalcitonin in some populations of patients has recently been reviewed.67,77 Boysen and colleagues67 assessed the diagnostic value of procalcitonin in post-operative infection. However, no conclusion could be drawn from their review because of significant heterogeneity among studies. Our analysis included one post-operative study,77 which was left out by this review. In another review, procalcitonin concentration was found to be better than C-reactive protein in diagnosing bacterial infection.77 However, this review included studies across a wide range of age groups, clinical settings, and disease spectrum. Additionally, nearly half of the study population (46%) included paediatric patients and many patients did not have SIRS (57%). Despite such a diverse case mix, the study did not assess heterogeneity or its effect on the pooled estimates, thus making it very difficult to interpret its findings.77 In view of these limitations, we applied in our study more strict inclusion criteria, focusing mainly on a more homogenous population, and used a substantially larger sample size (2092 vs 588). We also explored systematically the issue of heterogeneity by use of meta-regression and subgroup analysis. Furthermore, sensitivity analysis confirmed that our findings were robust and consistent. These methodological strengths have therefore enhanced the validity and applicability of our findings.

Publication bias is common in diagnostic studies and is possibly more of a problem than in studies of randomised controlled trials.79 We detected publication bias in our review. As expected, the missing studies were located to the left of the funnel plot, consistent with the general observation that studies with less optimistic estimation of diagnostic performance are less likely to get published. With imputed values, the re-calculated diagnostic OR was significantly lower than the observed value, indicating that the true diagnostic performance of procalcitonin could have been even lower. However, the statistical methods used to assess publication bias have limitations.79,84 The above findings therefore need to be interpreted in this context.

The scope of this review means that our findings cannot be generalised to specific diseases (eg, pancreatitis, burns) or settings (eg, cardiothoracic surgical patients, neonatal/paediatric patients). Our study did not include patients who were not critically ill, or who did not fulfil the SIRS criteria. The variation in disease prevalence and severity in these patients means that the diagnostic accuracy of procalcitonin is likely to be different, depending on the chosen population or setting. Finally, we did not include studies that assessed the ability of procalcitonin to diagnose septic shock, since these conditions were usually recognised by simple clinical criteria.

The focus of this review is on the role of procalcitonin in distinguishing sepsis from SIRS in critically ill patients. However, infection can be present without any clinical manifestation of SIRS.4 The role of procalcitonin in such a setting remains undefined, since most of the procalcitonin studies in this review used SIRS patients in the control groups. Furthermore, this review does not address the issue of prognosis. Further studies would be needed to assess the role of procalcitonin in both these settings.

Although the SIRS criteria are widely used in the literature surveyed by this review, they have been criticised for being too sensitive.69 However, this low threshold for detection is appropriate for a test for which the consequences of over-detection are outweighed by the consequences of undertreatment of potentially septic patients.81 Additionally, the SIRS criteria provide uniformity in inclusion criteria and allow valid comparison to be made across many different studies.84 Such uniformity has ensured the validity of statistical pooling in our meta-analysis. Despite its limitations, the continuing use of the SIRS concept has recently been supported by an international sepsis definitions conference.85 The findings of our study therefore reflect the prevalent use of the SIRS concept in sepsis research.79,86

Ideally, the additive value of the procalcitonin test to supplement a clinician’s bedside assessment should be evaluated in any diagnostic study. Unfortunately, most of the 18 studies did not explore how procalcitonin could be used to enhance clinical assessment, which highlights a recent trend of adopting a biomarker-based approach to diagnose sepsis. In light of our findings, future research should focus on incorporating biomarkers as part of an overall assessment of critically ill patients, rather than in preference to clinical assessment.

In summary, we found that procalcitonin had a low diagnostic performance in differentiating sepsis from SIRS in critically ill adult patients. The evidence presented in this review does not lend support to the widespread use of the procalcitonin test for sepsis diagnosis in critical care settings.

Conflicts of interest
We declare that we have no conflicts of interest.

Acknowledgments
We thank Marek Nalos, Joachim Brueckner, and Elri Jordaan for the translation of foreign language papers, and Ian Seppelt and Stephen Huang for their helpful comments on the manuscript. There was no funding source for this review. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

References


72 Jaeschke R, Guyatt G, Sackett D. User’s 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: 703–07.


