I, Sarah E Fitzgerald M.D., hereby submit this original work as part of the requirements for the degree of Master of Science in Clinical and Translational Research.

It is entitled:
A Meta-Analysis of the Diagnostic Performance of Procalcitonin in the Diagnosis of Serious Bacterial Infection in Pediatric Febrile Neutropenia

Student’s name: Sarah E Fitzgerald M.D.

This work and its defense approved by:

Committee chair: Erin Nicole Haynes, DrPH
Committee member: Paul Succop, PhD
Committee member: Jeffrey Welge, PhD
A Meta-Analysis of the Diagnostic Performance of Procalcitonin in the Diagnosis of Serious Bacterial Infection in Pediatric Febrile Neutropenia Episodes

A thesis submitted to the
Graduate School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Master of Science in Clinical & Translational Research

In the Department of Environmental Health
Division of Epidemiology & Biostatistics
of the College of Medicine

June, 2012

by

Sarah E. Fitzgerald

M.D, University of Cincinnati College of Medicine, June 2005
B.S., University of Dayton, May 2000

Committee Chair: Erin Haynes, DrPH
ABSTRACT

Background:
Continued advancements and intensification of chemotherapy regimens have enhanced outcomes for children with cancer. With more aggressive therapy, however, infection rates due to immune system suppression have increased\(^1\). Although progress in supportive care strategies has improved management of such infections\(^2\), infection remains a major cause of morbidity and mortality among pediatric oncology patients\(^3\). Serious bacterial infections often present non-specifically with only fever and neutropenia; however, infection represents only one etiology of fever and neutropenia.

Objective:
Identification of a biomarker to more quickly diagnose serious infection in fever and neutropenia episodes would be advantageous. Prior small studies suggest that procalcitonin may be a feasible and reliable marker for bacterial infection in febrile neutropenic pediatric patients. Because the implementation of such a biomarker would be significant to pediatric oncology practice, stronger evidence is required before the assay can be meaningfully, reliably, and safely applied to patient care.

Design/Methods:
A systematic review and meta-analysis of the literature regarding the use of procalcitonin as a marker of serious infection in neutropenic febrile children was performed. Several electronic databases that included records from 1966 to July 2011 were queried using the keyword “procalcitonin” alone and in combination with several search terms related to our topic.

Results:
The ten selected studies included 591 pediatric oncology patients with a total of 1161 febrile episodes. Regression analysis and the random effects model were used to estimate the common effect size for each measure of diagnostic performance for procalcitonin: pooled odds ratio was 39.29, with 95% confidence interval (12.61, 122.4); sensitivity was 0.709, or 70.9%, with p<0.0001 and 95% confidence interval (0.357, 1.06); specificity was 0.826, or 82.6%, with p<0.0001 and 95% confidence interval (0.629, 1.02).

Conclusions:

Procalcitonin is an appropriate candidate biomarker for serious bacterial infection in pediatric oncology fever and neutropenia patients, given that the odds of having an elevated procalcitonin assay and a serious bacterial infection are 39.29 times higher than the odds of having an elevated procalcitonin assay in the absence of serious bacterial infection. Furthermore, this meta-analysis produced evidence of modest to good sensitivity and better specificity with values of 71% and 83%, respectively.

References:


# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tables and Figures Summary</td>
<td>vi</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td></td>
</tr>
<tr>
<td>Search Strategy</td>
<td>3</td>
</tr>
<tr>
<td>Inclusion and Exclusion Criteria</td>
<td>4</td>
</tr>
<tr>
<td>Study Quality Assessment</td>
<td>6</td>
</tr>
<tr>
<td>Statistical Methods</td>
<td>8</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
</tr>
<tr>
<td>Search Results</td>
<td>10</td>
</tr>
<tr>
<td>Included Studies’ Characteristics</td>
<td>10</td>
</tr>
<tr>
<td>Model Results</td>
<td>16</td>
</tr>
<tr>
<td><strong>Discussion</strong></td>
<td>21</td>
</tr>
<tr>
<td><strong>References</strong></td>
<td>24</td>
</tr>
</tbody>
</table>
TABLES AND FIGURES

Table 1: Characteristics of the Included Studies

Table 2: Individual Study Results

Figure 1a: Forest Plot of Individual Study Results for Ln(OR) Analysis

Figure 1b: Forest Plot of Individual Study Results for Sensitivity Analysis

Figure 1c: Forest Plot of Individual Study Results for Specificity Analysis

Figure 2a: Funnel Plot for the Ln(OR) Analysis

Figure 2b: Funnel Plot for the Sensitivity Analysis

Figure 2c: Funnel Plot for the Specificity Analysis
**Introduction**

Continued advancements and intensification of chemotherapy regimens have enhanced outcomes for children with cancer. With more aggressive therapy, however, infection rates due to immune system suppression have increased\(^1\). Although progress in supportive care strategies has improved management of such infections\(^2\), infection remains a major cause of morbidity and mortality among pediatric oncology patients\(^3\).

Infections in this vulnerable population remain difficult to manage for numerous reasons, including the often non-specific clinical presentation of even serious bacterial infections. At infection onset, the patient may display only fever and neutropenia (absolute neutrophil count $<500$/$\text{mm}^3$), but then rapidly progress to septic shock and death within hours\(^4\). Because of the potential for such adverse outcomes, all pediatric oncology patients with fever and neutropenia are treated uniformly with rapid administration of broad-spectrum antibiotics and hospital admission. While this approach has improved survival from infection, neutropenic fevers can have less threatening etiologies, including pharmaceutical reactions, blood product exposure, viruses and the underlying malignancy. As a result, some pediatric fever and neutropenia patients are unnecessarily exposed to antibiotics and their potential adverse effects and to the psychosocial stressors of hospitalization\(^5\).

The current diagnostic gold standard for serious bacterial infection during febrile neutropenia episodes is a culture of blood or other body fluid that identifies an organism; however, this standard has flaws. Cultures may take hours to days to grow with additional time required for identification and susceptibility testing on the pathogen. Furthermore, the rapid administration of antibiotics can decrease the likelihood of obtaining a positive diagnostic blood culture. Finally, cultures are not 100 percent sensitive; frequently, serial blood cultures are required to identify the pathogen, and often a causative agent is never identified. For these reasons, identification of a serum biomarker to quicken the diagnosis of severe infection in pediatric febrile neutropenic patients would be advantageous. Characteristics of this biomarker
include being sensitive and specific for serious infection, detectable early in the infectious process, relatively inexpensive and feasible to obtain routinely, and reliable in pediatric neutropenia patients. In addition, an optimal biomarker would be able to identify both those patients at highest risk for serious bacterial infection and infectious complications who would require closer observation and prolonged inpatient management and those patients at lower risk for infection who may benefit from early discontinuation of antibiotic therapy and more rapid discharge from the inpatient setting.

Procalcitonin (PCT) has recently become a candidate biomarker. PCT, a precursor protein to the thyroid hormone calcitonin, was first reported as a marker for bacterial infections in 1993 by Assicot, et al. Studies have shown that PCT increases in serum just hours into the infectious process and begins to diminish after 48 hours if its stimulus ceases. This timing makes it advantageous over cytokines such as interleukin-6 that decrease rapidly (within 12 hours) following bacterial exposure, and over markers such as C-reactive protein that increase relatively late after endotoxin stimulation. While PCT can increase in response to other inflammatory processes such as trauma and malignancy, these increases are typically less than the rises in serum levels following bacterial endotoxin exposure. Studies of PCT levels conducted in emergency departments have demonstrated feasibility of performing the test. PCT has been shown to be a sensitive and specific marker for serious infection in a number of populations, including neonates, trauma patients, intensive care patients and adult oncology patients with fever and neutropenia.

Because of its promise for a role in the prediction of severe bacterial infection, PCT has gained attention in pediatric oncology research, and several studies have examined PCT as an early marker for bacterial infection in pediatric patients with fever and neutropenia. In general, studies to date have been small and single-center in nature. Although overall the results of these smaller studies suggest that PCT may be a feasible and reliable marker for bacterial infection in febrile neutropenic pediatric patients, and because the identification of such a biomarker would be
significant to pediatric oncology practice, stronger evidence is required before the assay can be
meaningfully, reliably and safely applied to patient care. We conducted a systematic review and
meta-analysis of the available data regarding the use of PCT as a marker of serious infection in
neutropenic febrile children.

Methods

Search Strategy

A literature search was conducted in April, 2011 using several databases, including the
National Institute of Health’s PubMed Database, which serves as the National Library of
Medicine (NLM®) journal literature search system with records from 1966 to April 2011, the
Cochrane Library and Cochrane Database of Systematic Reviews (copyright ©2011), and Google
Scholar. The search was repeated in July 2011 prior to data analysis for completeness. In each of
the databases, the keyword “procalcitonin” was used alone and in combination with the following
terms: pediatric, neutropenia, neutropenic, fever, febrile, sepsis, bacteremia, infection, oncology,
cancer and biomarker. Reference sections of included articles were also reviewed to supplement
our online database search. The search was limited to human studies in a clinical setting with a
minimum of an abstract available in the English language. Abstracts of the articles were used to
confirm the population and biomarker of interest. Once abstracts of potentially eligible studies
were identified, the full texts were carefully evaluated for inclusion into the meta-analysis. A
single investigator identified the eligible articles and determined the final included studies for the
analysis.
**Inclusion and Exclusion Criteria**

Because this is a meta-analysis of the utility of PCT as a biomarker of serious infection in neutropenic febrile children with cancer, included studies were limited to those with the target population of febrile neutropenic pediatric oncology patients. Although two studies included age maximums that are traditionally considered to be adult (31.8 years and 35.1 years), they were included in the meta-analysis as the care of these young adult cancer patients often falls to pediatric oncologists\textsuperscript{13}, especially when their diagnoses have characteristics of pediatric tumors and therefore benefit from pediatric oncology therapy protocols. Included studies were also limited to those conducted within large, tertiary pediatric oncology institutions to provide more uniformity regarding the patients treated, the chemotherapy treatment protocols administered and the management of febrile neutropenic episodes.

In addition, only studies that included patients with neutropenia secondary to chemotherapy treatment were included. Numerous prior studies have demonstrated the ability of neutropenic oncology patients to produce a procalcitonin response in the face of significant infection\textsuperscript{12,14,15,16}. Those that studied patients with neutropenia attributed to a new cancer diagnosis or an infection without recent chemotherapy administration were excluded, as it is unclear whether these conditions alter procalcitonin levels. Studies that included bone marrow transplant patients or patients with primary immune deficiencies were also excluded due to uncertain behavior of procalcitonin levels in these patients.

Included studies were prospective and observational in design. In order for a study to be included in the analysis, it had to obtain a procalcitonin level on its subjects within the first 24 hours of presentation of the febrile neutropenic episode. It was felt that a biomarker obtained within this time frame would represent a useful early marker in the recognition of serious bacterial infection; further, a defined sampling time permits a more refined description of sensitivity and specificity. In addition, the meta-analysis includes studies that administered antibiotics only after (or simultaneously with) the collection of blood cultures and initial
laboratory studies. A meaningful comparison of the biomarker procalcitonin with the diagnostic standard of a blood culture requires optimal application of the current standard, which in this case includes obtaining cultures without potential interference from pretreatment with antibiotics.

The diagnostic outcomes of interest for studies included in this meta-analysis varied slightly. Several studies\textsuperscript{17,18,19,20,21,22} examined procalcitonin as a diagnostic marker of bacteremia or sepsis, and one\textsuperscript{8} study analyzed procalcitonin as a marker of systemic bacterial infection. The remaining studies\textsuperscript{23,24,25} included in the analysis evaluated procalcitonin as a marker for any serious bacterial infection. Despite these minor variations, it was felt that studies with any of these three outcomes would be acceptable for inclusion. The goal of applying a biomarker such as procalcitonin in the clinical scenario of pediatric febrile neutropenia is to differentiate the patients at high risk of infection from those at low risk, those in whom continued hospital care is necessary from those who can be discharged, and those at high risk of complications from those at lower risk. In pediatric febrile neutropenia, the majority of bacterial infections requires at least initial inpatient care and close clinical monitoring due to the high risk nature of serious infection in immune compromised patients. For this reason, studies with any type of serious bacterial infection as the diagnostic endpoint were eligible for inclusion. Furthermore, only including studies with the most severe outcomes, such as gram negative bacteremia or sepsis, may inflate the estimated diagnostic accuracy of procalcitonin. By using a more general outcome of interest, the meta-analysis of procalcitonin’s accuracy may more closely reflect the assay’s true performance.

In order to be included in the study, the prevalence of the outcome in each study’s patient population had to be similar to the outcome’s prevalence as reported in current literature. While sensitivity and specificity have traditionally been described as being independent of the outcome’s prevalence, they may be influenced by qualities of the study’s population. According to the text by Szklo and Nieto\textsuperscript{28}, “[t]his is particularly true for conditions based on a continuous scale that is more or less arbitrarily changed into a binary one” (such as procalcitonin levels), and
the risk of introducing misclassification bias “tends to be higher for individuals whose true values are near the chosen cut-off value.” According to Santaloya, et al\textsuperscript{26}, about 25\% of blood cultures obtained in patients with clinical and laboratory evidence of an invasive bacterial infection will identify a pathogen. Similar findings are documented by Ammann, et al\textsuperscript{27}, who reported rates of 10-30\% for bacteremia and 20-55\% for any serious bacterial infection in pediatric febrile neutropenic episodes.

Lastly, in order to be part of the analysis, the included studies’ reported data had to include: the desired measures of effect, sensitivity and specificity; the cut-off threshold for procalcitonin values upon which sensitivity and specificity were based; the total number of febrile neutropenic episodes; the number of events (serious bacterial infections) documented in the study; either the actual values for or the data available to calculate the number of true positives, false negatives, true negatives, and false positives in each study. A standard form was used to extract the preceding data and other pertinent information from the included studies (Table 1). This data extraction was performed by one individual.

**Quality Assessment of Individual Studies**

In order to maximize the quality of the meta-analysis, steps were taken to ensure the validity of the included primary studies as described by Irwig, et al\textsuperscript{29}, who recommend that each study identify an acceptable diagnostic reference standard to which the test of interest is compared. In this meta-analysis, if the study endpoint was any significant bacterial infection\textsuperscript{8,23,24,25}, the study’s diagnostic criterion had to be based upon an established diagnostic guideline such as that established by the International Immunocompromised Host Society (ICHS)\textsuperscript{30}, whose requirements are as follows. Microbiologically documented infections require fever in addition to a positive culture from the blood, urine, throat or stool. Clinically documented infections require fever in addition to unambiguous clinical and/or radiographic evidence of infection, such as an evident pneumonia on chest X-ray or obvious soft tissue
infection. If the study endpoint was bacteremia\textsuperscript{17,19,20,21,22}, a positive blood culture in the presence of fever was the required diagnostic definition, again in line with ICHS standards. If the study’s outcome of interest was sepsis\textsuperscript{18,22}, the authors’ definition of sepsis needed to be based upon an accepted standard, such as that from the American College of Chest Physicians and the Critical Care Medicine Consensus statements\textsuperscript{31}. These professionals define sepsis as fever in addition to tachycardia and/or tachypnea; septic shock was defined as clinical sepsis with the necessity of inotropes or vasopressors to maintain hemodynamic stability despite adequate fluid resuscitation. As a corollary to this step, our meta-analysis also required that each author’s definitions of fever and neutropenia were consistent with standard definitions to improve the uniformity of the patient populations and clinical scenarios.

Irwig, et al\textsuperscript{29} also recommends that the test under investigation and the standard test be read independently, as lack of independence in test interpretation may over-estimate diagnostic accuracy. While not explicitly stated, this standard can be inferred in the included studies. The investigational test in each study is the serum procalcitonin level measured by immunoluminescence assay, while the diagnostic standards were dependent on microbiological findings of a positive body fluid culture, radiographic evidence of an infection or clinical signs and symptoms consistent with infection. In addition to being evaluated by specialists in unique departments, the tests are resulted on differing timetables; for example, procalcitonin assays require 1-2 hours, clinical diagnoses are based upon observation both immediately and over time, and cultures typically require hours to days of incubation.

Verification bias was reduced by methods recommended by Irwig, et al\textsuperscript{29} in order to more precisely define the experimental test’s accuracy. Verification bias can occur when the standard is applied only to those patients with positive experimental test results or only to those negatively resulted experimental tests from patients with clinical signs and symptoms suggestive of disease presence. In each included study, the diagnostic standard test and the experimental test of interest were both performed on every patient. In addition, studies were included only if the procalcitonin
assay and body fluid cultures were collected at the same point in time during the patient’s febrile episode. Further, each study assessed the accuracy of the experimental procalcitonin assay against an accepted diagnostic standard. According to Irwig, et al, this strengthens the analysis compared to a design that compares a set of primary studies that evaluates the standard test’s accuracy to a set of primary studies that evaluates the experimental test’s accuracy.

**Statistical Methods**

The performance of the procalcitonin assays in the individual studies was measured by sensitivity and specificity. The studies’ reports of sensitivity, specificity and total number of events were used to calculate the number of true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP). From these calculations, the true positive rate (TPR, or sensitivity) and the false positive rate (FPR, or 1 – specificity) and their respective logits, the odds ratio (OR) and its natural log (ln OR) and the variance (VAR) for each study were determined. A traditional Receiver Operating Characteristic (ROC) Curve is created by plotting pairs of true positive and false positive rates. However, the varying procalcitonin thresholds applied in the included studies represent a possible source of between-study heterogeneity that may cause traditional pooling of data to become inappropriate; simply averaging true and false positive rates from different studies may underestimate the test’s true performance. Therefore, the Summary Receiver Operating Characteristic (SROC) Curve method was applied in this analysis. This SROC method assumes that at least some of the variance in study results is due to the use of different thresholds and permits the combination of TPRs, FPRs, ORs, sensitivities and specificities obtained from studies that applied variable procalcitonin thresholds. The initial step in this method is the application of the following linear model:

\[ D = \alpha + \beta S. \]
In the model, the difference $D = \logit(\text{TPR}) - \logit(\text{FPR}) = \ln(\text{OR})$. The sum $S = \logit(\text{TPR}) + \logit(\text{FPR})$, $\alpha$ is the intercept, also an ln odds ratio, and $\beta$ is the regression coefficient of variable S, which describes the extent to which the odds ratio depends upon the procalcitonin threshold applied.

Least squares weighted and unweighted regression analyses were performed on the above linear equation. Using these results, it was determined that individual test measures of accuracy could be expressed as pooled weighted overall estimates for an odds ratio, sensitivity and specificity. Heterogeneity, or variation in the outcomes of the included studies, was examined. To assess this between-study heterogeneity, 95% confidence intervals were constructed for the individual studies, with significant heterogeneity being represented by multiple predicted values for either the log OR, sensitivity or specificity being outside these predicted intervals. This method of assessing heterogeneity was based upon that described by De Vries, et al\textsuperscript{33}. The traditional Cochran’s Q test of heterogeneity was also applied. The Q statistic has a chi-square distribution with $[k - 1]$ degrees of freedom, where $k$ is the number of studies. The applied pooling method was the random effects model, which permits effect sizes to vary from study to study as a function of both the individual studies’ characteristics and a random effect resulting from unidentified sources of heterogeneity of effect size between the studies\textsuperscript{34}. Regression methods and the random effects model were performed using SAS statistical software. As there was no missing data from our included studies, this was not a statistical issue the analysis addressed. A sensitivity analysis was performed by repeating the meta-analysis, each time omitting one study from the included studies to help describe the contribution of each individual study. Publication bias was also addressed in this meta-analysis.
Results

The Search

The original search yielded 164 electronic records: 60 from the PubMed database, 102 from Google Scholar, and 2 from the Cochrane Library. Review of abstracts led to the exclusion of 147 publications. The remaining 17 records were reviewed in full. Two studies were excluded because they included bone marrow transplant patients. One study was excluded because the authors examined change in procalcitonin levels in their patients, not raw procalcitonin values. One publication was excluded because the subjects were adult patients, which was not apparent from the abstract, and another was excluded because it did not report sensitivity, specificity or other desired measures of effect for procalcitonin. Another was excluded from the analysis because only 48% of its study population was neutropenic. The last excluded study was eliminated because its outcome of interest was not a serious bacterial infection. While this final publication did examine the use of procalcitonin to predict a favorable versus unfavorable outcome in pediatric febrile neutropenic episodes, unfavorable outcomes included either microbiologically or clinically diagnosed infections or persistent and/or recurrent fever; prolonged fever was not a diagnostic outcome this analysis addressed due to its often ambiguous etiology. In total, 10 publications were included in the meta-analysis. No additional studies were identified for inclusion from the follow-up search.

Characteristics of Included Studies

The ten selected studies included 591 pediatric oncology patients with a total of 1161 febrile episodes. A summary of the study characteristics is provided in Table 1. As previously mentioned, two studies\textsuperscript{19,20} that included patients of ages that are traditionally considered to be adult (31.8 years and 35.1 years) were included in the meta-analysis as young adult cancer
patients with “pediatric-like” diagnoses often receive medical care from pediatric oncologists. The included data was collected by pediatric oncology research from a geographically varied group of institutions including 2 studies each from Germany and Greece and 1 study each from Italy, Mexico, Netherlands, Slovenia, Turkey and the United States (Washington, D.C.). Three studies examined the diagnostic role of procalcitonin during febrile neutropenic episodes in pediatric acute lymphoblastic leukemia (ALL); the remaining 7 reports included patients with any pediatric oncology diagnosis.

Table 1: Characteristics of the Included Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of Patients</th>
<th>No. of Febrile Episodes</th>
<th>Age (Years)</th>
<th>Sex</th>
<th>Diagnosis</th>
<th>Outcome of Interest</th>
<th>Study Prevalence of Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hatzistilianou</td>
<td>17 2010</td>
<td>104</td>
<td>221</td>
<td>1-14</td>
<td>* Acute Lymphoblastic Leukemia</td>
<td>Bacteremia: Fever and positive blood culture</td>
<td>47.5%</td>
</tr>
<tr>
<td>Martinez-Albarran</td>
<td>2009</td>
<td>54</td>
<td>54</td>
<td>&lt;= 18</td>
<td>39% female Any Oncology Diagnosis</td>
<td>Systemic Bacterial Infection: Fever with positive blood or urine culture or clinical sepsis</td>
<td>33.3%</td>
</tr>
<tr>
<td>Stryjewski</td>
<td>2005</td>
<td>56</td>
<td>56</td>
<td>&lt;= 18</td>
<td>48% female Any Oncology Diagnosis</td>
<td>Sepsis: By modified American College of Chest Physicians/Society of Critical Care Med Consensus</td>
<td>28.6%</td>
</tr>
<tr>
<td>Fleischhack17</td>
<td>2000</td>
<td>51</td>
<td>122</td>
<td>0.7 – 31.8</td>
<td>39% female Any Oncology Diagnosis</td>
<td>Gram Negative Bacteremia</td>
<td>10.7%</td>
</tr>
<tr>
<td>Fleischhack18</td>
<td>2000</td>
<td>120</td>
<td>376</td>
<td>0.3 – 35.1</td>
<td>37% female Any Oncology Diagnosis</td>
<td>Gram Negative Bacteremia</td>
<td>5.9%</td>
</tr>
<tr>
<td>Hatzistilianou20</td>
<td>2007</td>
<td>29</td>
<td>94</td>
<td>1-14</td>
<td>* Acute Lymphoblastic Leukemia</td>
<td>Serious Bacterial Infection: Microbiologic or clinical diagnosis by ICHS Criteria</td>
<td>63.8%</td>
</tr>
<tr>
<td>Hitoglou-Hatz22</td>
<td>2005</td>
<td>67</td>
<td>67</td>
<td>1-14</td>
<td>63% female Acute Lymphoblastic Leukemia</td>
<td>Serious Bacterial Infection: Microbiologic or clinical diagnosis by ICHS Criteria</td>
<td>43.3%</td>
</tr>
<tr>
<td>Secmaer21</td>
<td>2007</td>
<td>49</td>
<td>60</td>
<td>2-18</td>
<td>41% female Any Oncology Diagnosis</td>
<td>Bacteremia: Fever and positive blood culture</td>
<td>10%</td>
</tr>
<tr>
<td>Kitanovski22</td>
<td>2006</td>
<td>32</td>
<td>68</td>
<td>1-18</td>
<td>34% female Any Oncology Diagnosis</td>
<td>Bacteremia or Sepsis: Fever and positive blood culture, +/- evidence of sepsis</td>
<td>23.5%</td>
</tr>
<tr>
<td>Miedema25</td>
<td>2010</td>
<td>29</td>
<td>43</td>
<td>Median: 8</td>
<td>49% female Any Oncology Diagnosis</td>
<td>Serious Bacterial Infection: Microbiologic or clinical diagnosis by ICHS Criteria</td>
<td>32.6%</td>
</tr>
</tbody>
</table>
All included studies applied a prospective and observational design, and they all obtained, at a minimum, a procalcitonin level within the first twenty-four hours of the patients’ febrile neutropenic presentation. In addition, 8 of the included studies performed serial procalcitonin measurements and provided variable accompanying degrees of analysis. This meta-analysis focuses on the ability of procalcitonin to enhance risk stratification and subsequent management of pediatric febrile neutropenic patients within 24 hours of their presentation; therefore, analysis of any serially drawn procalcitonin levels is not addressed here. Included studies varied only modestly regarding their outcome of interest. Six studies examined procalcitonin as a diagnostic marker of bacteremia or sepsis, one study analyzed procalcitonin as a marker of systemic bacterial infection, and the remaining three studies evaluated procalcitonin as a marker for any serious bacterial infection. In most study populations, the frequency of the outcome of interest reflects the prevalence reported in current literature. The two exceptions are Hatzistilianou, et al, which reports a 47.5% prevalence of bacteremia in their patient population, and Hatzistilianou, et al, which diagnosed a serious bacterial infection in 63.8% of its study patients. Every individual study provided sufficient data to calculate the required statistical measures for the meta-analysis: number of true positives, false negatives, true negatives and false positives; the TPR (sensitivity) and the FPR (1 – specificity) and their respective logits; OR; and variance. Individual study results are summarized in Table 2 and in the forest plots in Figures 1a, 1b and 1c, which were constructed from the worksheet developed by Clark and Djulbegovic.
Table 2: *Individual Study Results*

<table>
<thead>
<tr>
<th>Study</th>
<th>Events/Sample Size</th>
<th>PCT Threshold (ng/ml)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>OR (95% CI)</th>
<th>Variance OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hatzistilianou et al. 2010</td>
<td>105/221</td>
<td>2.0</td>
<td>0.940</td>
<td>0.965</td>
<td>431.95</td>
<td>0.424</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.871, 0.975)</td>
<td>(0.908, 0.987)</td>
<td></td>
<td>(120.53, 1548.01)</td>
<td></td>
</tr>
<tr>
<td>Martinez-Alberanz 2009</td>
<td>18/54</td>
<td>0.67</td>
<td>0.722</td>
<td>0.805</td>
<td>10.72</td>
<td>0.454</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.464, 0.893)</td>
<td>(0.634, 0.912)</td>
<td></td>
<td>(2.56, 40.15)</td>
<td></td>
</tr>
<tr>
<td>Stryjewski et al. 2005</td>
<td>16/56</td>
<td>0.5</td>
<td>0.940</td>
<td>0.900</td>
<td>141.00</td>
<td>1.386</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.680, 0.997)</td>
<td>(0.754, 0.967)</td>
<td></td>
<td>(14.03, 1416.84)</td>
<td></td>
</tr>
<tr>
<td>Fleischhacker et al. 2000</td>
<td>13/122</td>
<td>0.5</td>
<td>0.600</td>
<td>0.850</td>
<td>8.50</td>
<td>0.392</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.310, 0.839)</td>
<td>(0.766, 0.909)</td>
<td></td>
<td>(2.49, 29.02)</td>
<td></td>
</tr>
<tr>
<td>Fleischhacker et al. 2000</td>
<td>22/376</td>
<td>0.5</td>
<td>0.625</td>
<td>0.825</td>
<td>7.86</td>
<td>0.214</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.398, 0.811)</td>
<td>(0.780, 0.862)</td>
<td></td>
<td>(3.18, 19.43)</td>
<td></td>
</tr>
<tr>
<td>Hatzistilianou et al. 2007</td>
<td>60/94</td>
<td>2.0</td>
<td>0.965</td>
<td>0.970</td>
<td>891.48</td>
<td>1.504</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.872, 0.994)</td>
<td>(0.829, 0.998)</td>
<td></td>
<td>(80.56, 9864.69)</td>
<td></td>
</tr>
<tr>
<td>Hiotelou-Hatzis 2005</td>
<td>29/67</td>
<td>2.0*</td>
<td>0.950</td>
<td>0.930</td>
<td>252.43</td>
<td>1.130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.783, 0.994)</td>
<td>(0.787, 0.984)</td>
<td></td>
<td>(31.42, 2028.01)</td>
<td></td>
</tr>
<tr>
<td>Sezonetz et al. 2007</td>
<td>6/60</td>
<td>0.3</td>
<td>0.333</td>
<td>0.940</td>
<td>7.82</td>
<td>0.687</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.167, 0.548)</td>
<td>(0.791, 0.989)</td>
<td></td>
<td>(1.54, 39.69)</td>
<td></td>
</tr>
<tr>
<td>Klimowski et al. 2006</td>
<td>16/68</td>
<td>0.55</td>
<td>0.98</td>
<td>0.706</td>
<td>36.33</td>
<td>1.167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.678, 0.997)</td>
<td>(0.562, 0.820)</td>
<td></td>
<td>(4.37, 301.96)</td>
<td></td>
</tr>
<tr>
<td>Miedema 2010</td>
<td>14/43</td>
<td>0.25</td>
<td>0.790</td>
<td>0.770</td>
<td>12.59</td>
<td>0.625</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.492, 0.945)</td>
<td>(0.573, 0.898)</td>
<td></td>
<td>(2.67, 59.33)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1a: Forest Plot of Individual Study Results for Ln(OR) Analysis
Figure 1b: Forest Plot of Individual Study Results for Sensitivity Analysis
Figure 1c: Forest Plot of Individual Study Results for Specificity Analysis

Model Results

Unweighted and weighted least squares regression analyses were performed on the linear model \( D = \alpha + \beta S \), with \( D \) as the dependent variable and \( S \) as the independent variable. As described above, \( D \) is a ln(OR); \( S \) is the sum: logit(TPR) + logit(FPR); \( \alpha \) is the intercept, also a ln(OR); and \( \beta \) is the regression coefficient of \( S \), describing the extent to which the OR depends upon the procalcitonin threshold used. The unweighted regression analysis produced the following values: the intercept, or \( \alpha \), of 4.04 (SE 0.592, \( p \leq 0.0001 \)); regression coefficient \( \beta \) for the variable \( S \) of 0.557 (SE 0.428, \( p = 0.230 \)). The weighted regression analysis (weighted on the inverse variance of the ln(OR)) produced the following values: the intercept, or \( \alpha \), of 3.68 (SE 0.632, \( p \leq 0.0004 \)); regression coefficient \( \beta \) for the variable \( S \) of 0.630 (SE 0.479, \( p = 0.224 \)). Because the regression coefficient \( \beta \) was near zero and not statistically significant for either regression solution, it appears that for this analysis, the OR as a measure of effect size is not
significantly dependent upon the procalcitonin threshold applied. Therefore, it was determined that measures of individual test accuracies could be pooled and expressed as a weighted OR, sensitivity and specificity to describe procalcitonin’s diagnostic performance.

Between-study variation was assessed by the method described by De Vries, et al[33]. The asymptotic variances of the ln (OR) \[ \text{VAR (ln (OR))} = \frac{1}{TP} + \frac{1}{FN} + \frac{1}{TN} + \frac{1}{FP} \], of sensitivity \[ \text{VAR sensitivity} = \frac{1}{TP} + \frac{1}{FN} \], and of specificity \[ \text{VAR specificity} = \frac{1}{TN} + \frac{1}{FP} \] were used to calculate 95% confidence intervals for ln(OR), sensitivity and specificity, with the aid of SAS software. The predicted values of each variable from the regression model were compared to the calculated confidence intervals for that variable. Predicted values of the ln(OR) for 6 of the 10 included studies, predicted values of sensitivity for 4 of the included 10 studies, and predicted values of specificity for 9 of the included 10 studies were outside these confidence intervals. This is evidence that marked heterogeneity was present between studies[33]. The Cochran Q values for each measure marker of performance (ln (OR), sensitivity and specificity) were extremely small. This indicates statistically significant evidence to reject the null hypothesis, which states that all between study variability is due to chance and not true differences in results. Again, this points to the presence of heterogeneity among the included studies.

Because the regression analysis demonstrated that the OR, sensitivity, and specificity were not significantly affected by the selected procalcitonin threshold, and because between-study heterogeneity is present among the included studies, the random effects model was the selected method to estimate the common effect size for each measure of performance. The estimate for ln(OR) provided by this model was 3.671, with \( p \leq 0.0001 \) and a 95% confidence interval of (2.534, 4.807). When the log odds values were converted, the estimate for the pooled odds ratio was 39.29, with a 95% confidence interval of (12.61, 122.4). This means that the odds of an elevated or positive procalcitonin assay in pediatric febrile neutropenic patients with a
serious bacterial infection is 39.29 times higher than the odds of an elevated procalcitonin assay in febrile neutropenic pediatric oncology patients without a serious bacterial infection. The estimate for sensitivity was 0.709, or 70.9%, with \( p \leq 0.0001 \) and a 95% confidence interval of (0.357, 1.06). The estimate for specificity was 0.826, or 82.6%, with \( p \leq 0.0001 \) and a 95% confidence interval of (0.629, 1.02).

A sensitivity analysis was performed to evaluate the robustness of the results of the meta-analysis, applying a method based upon the description on pages 149-150 of Sutton, et al\textsuperscript{32}. This jackknife type of sensitivity analysis involves systematically repeating the analysis for each measure of performance, omitting a different study each time until each study has been excluded once. This helps both to identify which, if any, studies are most influential and to identify if the overall pooled effect relies heavily on the inclusion or exclusion of a single study. This process did not reveal strong effects from any single study for either the ln(OR), the sensitivity or the specificity analyses. For the ln(OR), the single study exclusion method for 4 studies resulted in ORs lower than that calculated when all 10 studies are included, and this method for the remaining 6 studies resulted in ORs higher than the 10-study pooled OR. The greatest deviation below the 10 study pooled OR was 0.348 on a log scale, and the greatest deviation above the 10 study pooled estimate was 0.218 on a log scale. For sensitivity, the single study exclusion method for 7 studies resulted in sensitivities lower than that calculated for all 10 studies, and this method for the remaining 3 studies resulted in sensitivities higher than the 10-study pooled sensitivity. The greatest deviation below the 10 study pooled sensitivity was 0.055, and the greatest deviation above the 10 study pooled estimate was 0.082. For specificity, the single study exclusion method for 6 studies resulted in specificities lower than that calculated when all 10 studies are included, and for the remaining 4 studies, specificities were higher than the 10-study specificity. The greatest deviation below the 10 study pooled specificity was 0.0057, and the greatest deviation above the 10 study pooled estimate was 0.015.
Funnel plots displaying the relationship between each measure of effect and its estimate of precision (1/standard error) were constructed to evaluate for the presence of publication bias (Figures 2a, 2b, 2c). As the ln(OR) plot is asymmetric about a ln(OR) of zero on the x-axis, it suggests the presence of a degree of publication bias. The funnel plot for the sensitivity data appears more randomly distributed, and the specificity data produces a plot shape similar to the traditional funnel plot. These findings indicate that publication bias may affect the latter two analyses to a lesser degree than that for the ln(OR). Certainly, however, with small study numbers, shapes of funnel plots should be interpreted with some caution. To further evaluate publication bias in this study, we used linear regression as described by Harling, et al\textsuperscript{36}, where the variable’s standard normal deviate (SND; calculated by either ((ln (OR)/SE), sensitivity/SE, or specificity/SE), is regressed against its precision (1/SE). The calculated intercept is a measure of symmetry where the greater the departure from zero, the greater the asymmetry present in the funnel plot. For ln(OR), the intercept was 0.427, p<0.0001, a significant difference from zero, which reinforces the prior finding that asymmetry is likely present in the ln(OR) funnel plot, and therefore an element of publication bias likely exists in its analysis. For sensitivity, the intercept was 0.829, p=0.145, which is not significantly different from zero for this regression. For specificity, the intercept was –0.125, p=0.424, again not significantly different from zero for these values. Compared to what was found in the funnel plots, these results are consistent in suggesting that publication bias may have less of an impact in the analyses for sensitivity and specificity than for ln(OR).
Discussion

This analysis, an example of the use of meta-analysis in the evaluation of the diagnostic performance of a clinical test, analyzed serum procalcitonin in the detection of serious bacterial infection in pediatric oncology febrile neutropenic episodes. Three measures of diagnostic performance were evaluated: OR, sensitivity and specificity. The calculated pooled OR was 39.29 with a 95% confidence interval of [12.61, 122.4], meaning that the odds of obtaining an elevated procalcitonin assay in pediatric febrile neutropenic patients with a serious bacterial infection is 39.29 times higher than the odds of finding an elevated procalcitonin assay in febrile neutropenic pediatric oncology patients without a serious bacterial infection, and that this result is statistically significant. The calculated pooled sensitivity was 70.9%, with a 95% CI of [0.357, 1.06], and the calculated pooled specificity was 82.6%, with a 95% CI of [0.629, 1.02]. As a degree of heterogeneity is likely present between the included studies, the random effects model was appropriately applied, which permits differences in effect sizes due to both known varying characteristics of the included studies and to a random effect from unidentified sources of variation. The sensitivity analysis indicates that the meta-analysis is robust in nature, as one study does not appear to dominate any of the pooled measures of effect.

The strengths and limitations of this analysis stem from both properties inherent to all meta-analyses and from features of this particular analysis. Like meta-analyses of clinical trials, the process of meta-analysis of diagnostic tests seeks to produce a more precise and more robust estimate of the true test accuracy by appropriately combining data from a set of smaller studies that evaluate the same test in similar clinical settings. Through this process, this analysis has used data from 10 smaller, single institution studies to produce an estimate now based upon 591 pediatric oncology patients with 1161 febrile neutropenic episodes. Although the individual included studies were conducted in similar clinical settings, there is still likely to be between-study variations present that produce differing effect sizes; this variation was appropriately addressed with the use of the random effects model for the data analysis.
Despite rigorous attempts to identify all the relevant studies and to conduct the most appropriate methods of data analysis, limitations remain. First, the individual studies are small in size with subsequently wide confidence intervals. In addition, the meta-analysis was conducted on a relatively small number of studies, which in turn produces estimates, although statistically significant, that have accompanying somewhat wide confidence intervals. Limited number of subjects per study and limited availability of pertinent studies is an issue not unfamiliar to pediatric oncology given the relative rarity of it as a disease. Instead of regarding the available data as unreliable, one must take this into account when applying this information to one’s own individual patients. There is evidence, however, of the robustness and generalizability of this analysis, as the sensitivity analysis did not reveal strong dominance of effect by any one individual study.

From the funnel plot and linear regression exercises, it appears that an element of publication bias is present in our analysis for ln(OR) but less so for the sensitivity and specificity analyses despite the best efforts for a complete and thorough literature search. Publication bias results from the probability that studies with positive or significant results are more likely to be published and therefore available for analysis than those studies that confirm a null hypothesis. The concern is that this bias will produce an over-estimate of the true effect size. In order to minimize this bias, efforts were taken to identify all potentially eligible studies. The literature search was extensive and conducted on three large scholarly databases. An initial database search was performed and then repeated prior to beginning the statistical analysis. In addition, the reference sections of all included studies and all background and methodology articles were reviewed for potential studies. Included studies represent data from large pediatric oncology institutions in eight countries. Efforts were taken to reduce bias of the final estimate with each step of this meta-analysis. Inclusion and exclusion criteria and the quality assessment methods were applied with much consideration. Further, the diagnostic outcome of interest for this meta-analysis was conservative. By evaluating the diagnostic accuracy of procalcitonin in the
diagnosis of any serious bacterial infection, rather than its performance in diagnosing only the most severe outcomes, a more conservative effect size was obtained.

From this meta-analysis, one can conclude that serum procalcitonin is an appropriate candidate biomarker for serious bacterial infection in pediatric oncology fever and neutropenia patients, given that the odds of having an elevated procalcitonin assay and a serious bacterial infection are 39.29 times higher than the odds of having an elevated procalcitonin assay in the absence of serious bacterial infection. In addition, this meta-analysis produced evidence of modest to good sensitivity and better specificity with values of 71% and 83%, respectively. These numbers suggest that procalcitonin may be best applied as a tool to help rule in serious bacterial infection if serum procalcitonin is elevated; knowledge that a patient is at high risk for infection and complications early in a patient’s febrile neutropenic course would be quite valuable. However, further research is required. The appropriate procalcitonin threshold value to discriminate between a normal and elevated assay is not yet known, and this will require further clinical studies to define. Once an appropriate diagnostic threshold is established, prospective clinical trials are necessary to evaluate the assay’s ability to identify both those patients at highest risk for serious bacterial infection and infectious complications who would require closer observation and prolonged inpatient management and those patients at lower risk for infection who may benefit from early discontinuation of antibiotic therapy and more rapid discharge from inpatient care. In addition, evaluation of serial procalcitonin measurements to predict outcomes in febrile cancer patients is another future research pursuit that can further aid the appropriate management of this common yet potentially tenuous clinical scenario in immune compromised pediatric patients with fever.
References


