Prediction Models for Neonatal Health Care–Associated Sepsis: A Meta-analysis

Evelien Hilde Verstraete, RN, MNSc, Koen Blot, BSc, Ludo Mahieu, MD, PhD, Dirk Vogelaers, MD, PhD, Stijn Blot, PhD

abstract

BACKGROUND AND OBJECTIVES: Blood culture is the gold standard to diagnose bloodstream infection but is usually time-consuming. Prediction models aim to facilitate early preliminary diagnosis and treatment. We systematically reviewed prediction models for health care–associated bloodstream infection (HABSI) in neonates, identified superior models, and pooled clinical predictors. Data sources: LibHub, PubMed, and Web of Science.

METHODS: The studies included designed prediction models for laboratory-confirmed HABSI or sepsis. The target population was a consecutive series of neonates with suspicion of sepsis hospitalized for ≥48 hours. Clinical predictors had to be recorded at time of or before culturing. Methodologic quality of the studies was assessed. Data extracted included population characteristics, total suspected and laboratory-confirmed episodes and definition, clinical parameter definitions and odds ratios, and diagnostic accuracy parameters.

RESULTS: The systematic search revealed 9 articles with 12 prediction models representing 1295 suspected and 434 laboratory-confirmed sepsis episodes. Models exhibit moderate-good methodologic quality, large pretest probability range, and insufficient diagnostic accuracy. Random effects meta-analysis showed that lethargy, pallor/mottling, total parenteral nutrition, lipid infusion, and postnatal corticosteroids were predictive for HABSI. Post hoc analysis with low-gestational-age neonates demonstrated that apnea/bradycardia, lethargy, pallor/mottling, and poor peripheral perfusion were predictive for HABSI. Limitations include clinical and statistical heterogeneity.

CONCLUSIONS: Prediction models should be considered as guidance rather than an absolute indicator because they all have limited diagnostic accuracy. Lethargy and pallor and/or mottling for all neonates as well as apnea and/or bradycardia and poor peripheral perfusion for very low birth weight neonates are the most powerful clinical signs. However, the clinical context of the neonate should always be considered.

www.pediatrics.org/cgi/doi/10.1542/peds.2014-3226

DOI: 10.1542/peds.2014-3226

Accepted for publication Jan 27, 2015

Address of correspondence Evelien Verstraete, RN, MNSc, Department of Internal Medicine, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium. E-mail: evelienh.verstraete@ugent.be
Models for HABSI in hospitalized neonates and to evaluate their diagnostic accuracy to find superior models. The second aim was to pool odds ratios (ORs) of individual clinical parameters to detect superior clinical predictors of HABSI in neonates.

METHODS
A systematic literature review, diagnostic accuracy assessment of the prediction models, and random effects meta-analysis of clinical parameters were performed. The results are reported in accordance with the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analysis).

Literature Search Strategy
A systematic search was done by 2 independent researchers (E.V., S.B.) on PubMed, LibHub, and Web of Science without language and time period restrictions (up to April 2014). After record screening based on abstract, language restrictions were applied. All keywords or Mesh terms were applied for [Title and/or Abstract] or [Topic] and contained four parts: (1) “bloodstream infection” or “bloodstream infection” or “sepsis” or “septic(a)emia” or “bacter(a)emia”; (2) “prediction model” or “diagnostic screening model” or “prediction score” or “screening score” or “diagnostic score” or “diagnostic markers” or “diagnostic tool” or “clinical markers” or “clinical signs” or “clinical characteristics” or “predictors”; (3) “neonate” or “newborn” or “preterm” or “neonatal” or “NICU”; (4) “healthcare(-) associated” or “hospital(-) acquired” or “nosocomial or late(-)onset.

Additional searches were performed by reviewing the bibliography of the retrieved full text articles and by a manual search of expert authors.

Study Selection Criteria
We included studies creating a prediction model for laboratory-confirmed HABSI with clinical signs. Laboratory-confirmed HABSI or sepsis was defined as a positive culture of blood, cerebrospinal fluid, pleural fluid, and/or urine; culturing was at least 48 hours after birth or admission. Target population was a consecutive series of hospitalized neonates with suspicion of sepsis. It was required that in these studies, relationships between clinical parameters and sepsis were assessed by univariate analyses, whereas prediction models were developed by regression modeling; sensitivities, specificities, or ORs needed to be reported. Clinical parameters under research must be recorded preceding or at time of culturing. The final prediction model needed to include at least 3 predictors; because neonatal sepsis is a complex clinical syndrome, prediction models based on 2 possible parameters may not be justified. A checklist with all inclusion criteria was used to assess eligibility of the studies, and when in doubt, issues for inclusion or exclusion were discussed between the 2 independent researchers.

The following items were collected: population characteristics, setting, methods, statistical methods, exclusion criteria, applied case definition, definition for suspected HABSI/sepsis, total suspected episodes, total HABSI/sepsis episodes, assessment time of a suspected episode, clinical parameter definitions, predictor/event ratio, and prediction model accuracy. Model diagnostic accuracy was assessed by pre- and posttest probabilities, sensitivities, specificities, and positive and negative likelihood ratios.

Quality Assessment
The Quality Assessment of Diagnostic Accuracy Studies–2 scale was used to evaluate the methodologic quality of the included studies.22 This validated scale assesses criteria on 4 domains: patient selection (consecutive or random sampling, inappropriate
exclusions, description of included patients), index test (description, execution, blinded interpretation), reference standard (accuracy, blinded interpretation), and flow and timing (application of and interval between reference standard and index test, exclusions for analysis). Because of limitations in quality scales, we also examined other individual components of methodologic standards for prediction models: clearly defined outcome and predictive variables, description of patient characteristics, and crossvalidation. In addition, overfitting, which is a substantial threat in multivariate regression modeling, was assessed by calculation of the number of predictors included in the regression model per total events; this is termed the predictor/event ratio.

**Statistical Methods**

Comparisons between groups were performed by using the Mann-Whitney U test for continuous variables and the Fischer exact test for categorical variables. The statistical package of social science version 21 was used for these tests (SPSS Inc, Chicago, IL).

A random effects meta-analysis using the inverse-variance method obtained ORs and 95% confidence intervals (CIs) for clinical predictors. This was performed by using the Comprehensive Meta-Analysis version 2 program (Biostat Inc, Englewood, NJ). Statistical heterogeneity was predefined by using the Higgins I² statistic (I² \( \leq 25\% \) for low, 25\% < \( I^2 \) < 50\% for moderate, and \( I^2 \geq 50\% \) for high). Only those parameters that were similar in definition were pooled to avoid clinical heterogeneity. A 100\% interobserver agreement (by E.V., K.B., S.B.) concerning these parameter definitions was required to be included in the random effects meta-analysis. Model diagnostic accuracy variables were calculated with a Web-based diagnostic test calculator.

Funnel plots will be constructed for assessment of publication bias when at least 10 studies are included.

**RESULTS**

In total, 80 studies were retrieved of which 9 were included, representing 12 prediction models or scores, a total of 1295 suspected and 434 laboratory-confirmed HABSI episodes. Figure 1 shows the results of the search strategy. The research period spanned from 1993 to 2007 with 3 western European, 4 South Asian, 1 Canadian, and 1 Turkish study. All researches were performed in level III settings of which one was conducted in a low-resource level III hospital.

**Excluded Studies Not Meeting Inclusion Criteria**

One study was excluded because the time frame of recorded clinical parameters encompassed a 24-hour follow-up postsepsis onset. Six studies developed a prediction model including early-onset or community-acquired episodes of sepsis. In several studies by Griffin et al., heart rate characteristics and clinical parameters are under study, and prediction models are developed. However, the studies of Griffin and colleagues as well as the research of Modi et al. considered all NICU patients rather than neonates with suspicion of HABSI.

**Methodologic Quality of the Included Studies**

The methodologic performance of the included studies could be considered as low-medium risk for bias. Patient selection was mostly well described, as was the motive for patient exclusion. Selected patients were all under suspicion of sepsis and underwent the reference test (ie, blood, urine, or cerebrospinal fluid culturing) and an index test (ie, clinical prediction score or model). Index tests were overall well developed using similar statistical methods. The reference test blood culturing is considered medium risk for bias because it is known not to be 100\% accurate. Because blood culture results are available only after at least 12 hours and the index test was performed before the blood culture test, the studies can be considered as double blinded, except for the 2 retrospective studies. Assessment of methodologic standards for prediction models exhibited clearly defined outcome for all studies, moderate description of patient characteristics, and occasionally clearly defined clinical parameters. In particular, the definitions of the clinical signs and the interobserver variability in interpretation might be a medium risk for bias. Two studies had no validation cohort but used a bootstrapping statistical technique for internal validation. Most studies defined the time of assessment as “day of sepsis workup,” so a time frame for data collection of a maximum of 24 hours preceding sepsis workup is acknowledged. The 2006 retrospective study of Dalgic et al. did not describe the applicable time frame for data collection. Concerning the issue of overfitting, the 2005 study of Okascharoen et al. overruled the generally accepted 1:10 predictor/event ratio, so this might be a risk for bias. A visual presentation rating the risk of bias in low, medium, and high is presented in Supplemental Appendix 1.

**General Description of the Included Studies**

Characteristics of the 9 included articles are shown in Table 1. Pretest probability of sepsis ranged between 17\% to 55\%, indicating an important variation in study population. Three studies did not include all neonates with suspicion of sepsis in their research; selection was made for neonates with gestational age.
<34 weeks$^{29,33}$ and for birth weight $\geq 1000$ g and $\leq 2500$ g$^{32}$ One study$^{35}$ did not report information on patient characteristics but did focus on nosocomial sepsis in neonates on neonatal intensive care. Two studies were internal$^{34}$ and external validation studies$^{28}$ of their former developed prediction score study. Another 2 studies$^{32,33}$ are adapted external validation studies of Singh et al.$$^{30}$

**Characteristics and Diagnostic Accuracy of the Prediction Models**

Characteristics and diagnostic accuracy of the 12 prediction models with several cutoff scores are tabulated in Table 2. It is observed that 3 models with a particular cutoff score have a sensitivity of at least 95%. Of these 3 models, Mahieu et al’s nosocomial sepsis score (NOSEP)$^{27}$ with a point score of 8 or higher, displays the highest specificity and positive likelihood ratio. Figure 2 is a visual comparison of the sensitivity and specificity of 7 models. If a model has several cutoff scores, it is represented with its highest sensitivity model.

**Random Effects Meta-analysis**

Odds ratios of clinical parameters in univariate analysis of 5 studies$^{27,29-31,33}$ were included in the random effects meta-analysis. Ten of the 29 clinical signs and 8 of the 22 risk factors within 2 to 5 studies could be pooled. Eleven clinical signs and 12 risk factors were research variables in 1 study; another 6 clinical signs and 2 risk factors could not be pooled due to heterogeneity in variable definition. Definitions of the pooled and nonpooled clinical signs and risk factors are presented in Supplemental Appendix 2. Cutoff values of biological markers were largely different, and data on laboratory techniques used were not available, so biological markers were not pooled. Pooled OR and statistical heterogeneity of all 18 individual clinical parameters are tabulated in Table 3. Figure 3 displays the forest plots of the 5 significant clinical parameters predictive for HABSI in neonates. Because 10 of 18 clinical parameters exposed medium to high statistical heterogeneity, a post hoc analysis was performed with 2 studies$^{29,33}$ including solely low gestational age neonates (<34 weeks) and 9 clinical signs. Pooled OR and statistical heterogeneity of these 9 signs are tabulated in Table 4. Figure 4 displays the forest plots of the 4 significant clinical signs predictive for HABSI in low gestational age neonates (<34 weeks).

**DISCUSSION**

To our knowledge, this is the first systematic review of clinical prediction models for HABSI in neonates. Three prediction models
<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Method</th>
<th>Setting</th>
<th>No. of Episodes (Prevalence)</th>
<th>Confirmed Episode Definition</th>
<th>No. of Signs Evaluated</th>
<th>No. of Predictors/Events</th>
<th>Exclusion Criteria</th>
<th>Validation Cohort</th>
<th>No. of Episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mahieu27 (2000)</td>
<td>Pro</td>
<td>Level III NICU, Belgium</td>
<td>Suspected: 104, confirmed: 43 (41%)</td>
<td>Positive BC. Extra criteria for BC with skin commensal: 2 positive BC from 2 venipunctures or ≥1 from CVC</td>
<td>17 CS, 20 BM, 12 RF</td>
<td>5/104</td>
<td>BC not drawn before starting antibiotics</td>
<td>Retro Internal</td>
<td>50</td>
</tr>
<tr>
<td>Mahieu28 (2002)</td>
<td>Pro</td>
<td>6 Level III NICUs, Belgium</td>
<td>Suspected: 93, confirmed: 51 (55%)</td>
<td>Positive BC. Extra criteria for BC with skin commensal: 2 positive BC from 2 venipunctures or ≥1 from CVC</td>
<td>None</td>
<td>No regression modeling</td>
<td>BC not drawn before starting antibiotics</td>
<td>Internal and external validation study of Mahieu et al. and testing of NOSEP-new-I and NOSEP-new-II, which are prospectively validated in an external cohort of 93 suspected episodes</td>
<td>Pro</td>
</tr>
<tr>
<td>Singh30 (2003)</td>
<td>Pro</td>
<td>Level III NICU, India</td>
<td>Suspected: 105, confirmed: 30 (29%)</td>
<td>Positive BC or CSF culture</td>
<td>16 CS</td>
<td>7/105</td>
<td>Major CM, &gt;28 days of life</td>
<td>Pro Internal</td>
<td>20</td>
</tr>
<tr>
<td>Okascharoen31 (2005)</td>
<td>Retro</td>
<td>Level III NICU plus regular and special care unit, Thailand</td>
<td>Suspected: 100, confirmed: 17 (17%)</td>
<td>Bacteria in blood, CSF, pleural fluid, bone joint, or urine, &gt;10^6 CFU/mL; extra criteria for BC with skin commensal: 2 positive BC from 2 punctures on the same day</td>
<td>8 CS, 5 BM, 17 RF</td>
<td>20/100</td>
<td>Life-threatening CM; suspicion of late-onset sepsis started within 7 d after ceasing antibiotics for early-onset sepsis; a second or more episode</td>
<td>Retro Internal</td>
<td>73</td>
</tr>
<tr>
<td>Daigic35 (2006)</td>
<td>Retro, matched cohort</td>
<td>Level III NICU, Turkey</td>
<td>Suspected: 132, confirmed: 51 (38%)</td>
<td>Positive BC</td>
<td>None</td>
<td>No regression modeling</td>
<td>No data</td>
<td>External validation of Mahieu et al. and testing of a clinical score system defined by the team</td>
<td>Pro Internal</td>
</tr>
<tr>
<td>Okascharoen34 (2007)</td>
<td>Pro</td>
<td>Level III NICU, Canada</td>
<td>Suspected: 105, confirmed: 35 (33%)</td>
<td>Positive BC or CSF culture; contamination is defined as ignoring a positive BC by the attending clinician and discontinuation of antibiotic therapy</td>
<td>None</td>
<td>No regression modeling</td>
<td>BC not drawn before starting antibiotics; previously discharged from the hospital; &gt;28 d of life</td>
<td>External validation of Okascharoen et al.</td>
<td>Pro</td>
</tr>
<tr>
<td>Kudawla32 (2008)</td>
<td>Pro</td>
<td>Level III NICU, India</td>
<td>Suspected: 220, confirmed: 80 (27%)</td>
<td>Clinical suspicion of sepsis with a positive BC</td>
<td>None</td>
<td>No regression modeling</td>
<td>Major CM, died ≤24 h after onset of illness</td>
<td>External validation of Singh et al. and testing a combined model with sepsis screen and clinical score with no internal or external validation</td>
<td>Pro</td>
</tr>
<tr>
<td>Rosenberg33 (2010)</td>
<td>Pro</td>
<td>Level III special care neonatal unit, Bangladesh</td>
<td>Suspected: 193, confirmed: 105 (54%)</td>
<td>Positive noncontaminated BC</td>
<td>21 CS</td>
<td>21/193</td>
<td>Major CM, &gt;72 h postnatal age before admission BC drawn &lt;4 d apart in same infant &gt;28 d of life</td>
<td>External validation of Singh et al. and development of improved score with no internal or external validation cohort</td>
<td>Pro</td>
</tr>
<tr>
<td>Bekhofer29 (2013)</td>
<td>Pro</td>
<td>Level III NICU, the Netherlands</td>
<td>Suspected: 187, confirmed: 50 (27%)</td>
<td>Positive noncontaminated BC with skin commensal: 2 positive BCs and/or signs of catheter-related bloodstream infection, ie, inflammation on skin at site of insertion</td>
<td>14 CS, 7 RF</td>
<td>6/187</td>
<td>Antibiotics 24 h before assessment time</td>
<td>No validation cohort</td>
<td>20</td>
</tr>
</tbody>
</table>

BC, blood culture; BM, biological markers; CM, congenital malformation; CS, clinical signs; CSF, cerebrospinal fluid; CVC, central vascular catheter; Pro, prospective; Retro, retrospective; RF, risk factors.
<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Model Name</th>
<th>Model Variables</th>
<th>Score</th>
<th>Model Application</th>
<th>Pre-TP (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Post-TP+</th>
<th>Post-TP-</th>
<th>LR+</th>
<th>LR-</th>
<th>Area Under the ROC Curve (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mahieu27 (2000) NOSEP 1, NOSEP 2</td>
<td>CRP ≥14 mg/L, Neutrophil fraction &gt;50%, Temperature &gt;38.2°C, TPN ≥14 d, Thrombocytopenia &lt;150 × 10^9/μL, NOSEP 1+, positive hub and exit site cultures</td>
<td>5</td>
<td>Score of ≥11+ positive culture</td>
<td>14</td>
<td>8</td>
<td>54</td>
<td>8</td>
<td>1.67</td>
<td>0.12</td>
<td>0.82 (0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahieu28 (2002) NOSEP-new-I</td>
<td>CRP ≥30 mg/L, Neutrophil fraction &gt;63%, Temperature &gt;38.1°C, TPN ≥15 d, Thrombocytopenia &lt;150 × 10^9/μL, NOSEP-new-I+ recent surgery+ maternal hypertension + ventilation</td>
<td>5</td>
<td>Score of ≥11+ 3 risk factors</td>
<td>11</td>
<td>11</td>
<td>41</td>
<td>26</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Singh30 (2003) Weighted clinical score for definite sepsis</td>
<td>Abdominal distention, Chest retraction, Grunting, Hyperthermia, Increased aspirates, Lethargy, Tachycardia</td>
<td>2</td>
<td>Score</td>
<td>No data</td>
<td>1</td>
<td>2</td>
<td>29</td>
<td>87</td>
<td>29</td>
<td>33</td>
<td>16</td>
<td>1.23</td>
</tr>
<tr>
<td>Okascharoen31 (2005) Bedside prediction score</td>
<td>Abnormal body temperature, Hypotension, Neutrophil bandemia ≥1%, Respiratory insufficiency, Thrombocytopenia &lt;150 × 10^9/μL, Umbilical line ≤7 d, Umbilical line &gt;7 d</td>
<td>3</td>
<td>Score</td>
<td>No data</td>
<td>4</td>
<td>5</td>
<td>17</td>
<td>82</td>
<td>74</td>
<td>39</td>
<td>5</td>
<td>3.15</td>
</tr>
<tr>
<td>Daigle35 (2005) Clinical Score</td>
<td>Abdominal distention, Bradycardia, Feeding intolerance, Hypotension, Lowest-highest body temperature difference, Respiratory symptom</td>
<td>2</td>
<td>Score</td>
<td>No data</td>
<td>6–12</td>
<td>39</td>
<td>56</td>
<td>71</td>
<td>55</td>
<td>28</td>
<td>1.35</td>
<td>0.62</td>
</tr>
<tr>
<td>Kudawla32 (2008) Clinical score</td>
<td>Abdominal distention, chest retractions, grunting, hyperthermia, lethargy, tachycardia, pre feed aspirates</td>
<td>No score</td>
<td>≥1 clinical sign</td>
<td>27</td>
<td>90</td>
<td>23</td>
<td>30</td>
<td>14</td>
<td>1.17</td>
<td>0.43</td>
<td>No data</td>
<td></td>
</tr>
<tr>
<td>Sepsis screen</td>
<td>Absolute neutrophil count; CRP; microerythrocyte sedimentation rate; neutrophil ratio (immature/total)</td>
<td>No score</td>
<td>≥2 clinical signs</td>
<td>27</td>
<td>52</td>
<td>65</td>
<td>36</td>
<td>21</td>
<td>1.49</td>
<td>0.74</td>
<td>No data</td>
<td></td>
</tr>
</tbody>
</table>
Lethargy and pallor and/or mottling appear to be the best predictive clinical signs, whereas the use of total parenteral nutrition (TPN), lipid infusion, and postnatal corticosteroids are significant iatrogenic risk factors that should strengthen suspicion for HABSI when clinical signs occur.

The strengths of this study include methodologic quality and diagnostic accuracy evaluation of the models and a random effects meta-analysis design.

### Diagnostic Accuracy of the Prediction Models

The variables of the included prediction models were largely different, so diagnostic test accuracy meta-analysis had no clinical connotation and was not executed. Considering diagnostic accuracy, pre- and posttest probability, sensitivity, specificity, and positive and negative likelihood ratios were assessed. First, a large range in pretest probability was noted, reflecting the fundamental variation in clinical condition of the study population and influencing the posttest probability and pretest probability differences. Second, diagnostic performance is a major concern. It is observed that only 3 models have sufficient sensitivity. A sensitivity of 95% is required for a potentially lethal condition such as HABSI. The NOSEP score of 8 of Mahieu et al.\(^27\) and CRP, C-reactive protein; LR+, or LR−, likelihood ratio of a positive or negative test; ROC, receiver operating characteristic; TP, test probability (+ or −, of a positive or negative test).

<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Model Name</th>
<th>Model Variables</th>
<th>Score</th>
<th>Model Application</th>
<th>Pre-TP (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Post-TP+</th>
<th>Post-TP−</th>
<th>LR+</th>
<th>LR−</th>
<th>Area Under the ROC Curve (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosenberg(^55) (2010)</td>
<td>Clinical sepsis risk score</td>
<td>Apnea, hepatomegaly, jaundice, lethargy, pallor</td>
<td>No score</td>
<td>≥2 markers</td>
<td>1 clinical signs</td>
<td>54</td>
<td>77</td>
<td>50</td>
<td>64</td>
<td>35</td>
<td>1.54</td>
<td>0.46</td>
</tr>
<tr>
<td>Behkof(^29) (2013)</td>
<td>Reduced model for proven sepsis</td>
<td>Capillary refill &gt;2 s, central line in past 24 h, increased respiratory support, lethargy</td>
<td>No score</td>
<td>≥2 clinical signs</td>
<td>1 of the 4 signs</td>
<td>54</td>
<td>42</td>
<td>82</td>
<td>73</td>
<td>45</td>
<td>2.33</td>
<td>0.71</td>
</tr>
</tbody>
</table>

CRP, C-reactive protein; LR+, or LR−, likelihood ratio of a positive or negative test; ROC, receiver operating characteristic; TP, test probability (+ or −, of a positive or negative test).

The diagnostic scores of the included prediction models were largely different, so diagnostic test accuracy meta-analysis had no clinical connotation and was not executed. Considering diagnostic accuracy, pre- and posttest probability, sensitivity, specificity, and positive and negative likelihood ratios were assessed. First, a large range in pretest probability was noted, reflecting the fundamental variation in clinical condition of the study population and influencing the posttest probability and pretest probability differences. Second, diagnostic performance is a major concern. It is observed that only 3 models have sufficient sensitivity. A sensitivity of 95% is required for a potentially lethal condition such as HABSI. The NOSEP score of 8 of Mahieu et al.,\(^27\) and CRP, C-reactive protein; LR+, or LR−, likelihood ratio of a positive or negative test; ROC, receiver operating characteristic; TP, test probability (+ or −, of a positive or negative test).
screen (2 biomarkers) of Kudawla et al., and the reduced model (1 sign) of Bekhof et al. exhibit 5% or fewer false-negative cases. In contrast, the specificity of these 3 models is low (range, 18%–43%); of these 3, the NOSEP score of 8 reports the highest specificity (43%), nonetheless indicating 33.7% false-positive cases (n = 35) and only a 1.67 chance on a positive test for cases versus noncases. However, when clinical deterioration is a criterion for culturing, those 33.7% false-positive cases might be considered true-positive cases for another illness, justifying antimicrobial therapy. In contrast, when a C-reactive protein of 1 mg/dL is a criteria for culturing and a NOSEP score of 8 is detected, the clinical condition of the neonate might not yet be taken into consideration; for example, a neonate receiving TPN $14 days with a neutrophil fraction $50% has a NOSEP score of 9. In this case, the score can indicate close monitoring of the neonate and increased or more frequent observation of clinical signs. Clinically, antibiotic treatment might be postponed until clinical deterioration or a score of 11 is reached so that overtreatment can be repressed. Overall, prediction models have been developed to streamline the plethora of signs and risks for HABSI and thus to facilitate medical judgment and decision-making concerning treatment. In clinical practice, the interpretation of the nonspecific clinical signs is pivotal, although not always obvious and subject to interobserver variability. The consistency with other symptoms as well as underlying conditions must be considered. Important here is the weight assigned to a clinical observation. It is not only the presence of a clinical sign but primarily a change in the presentation of that clinical sign that might lead to accurate prediction of HABSI. For example, premature neonates can have more respiratory distress than full-term neonates, so signs such as apnea, chest retraction, and grunting might be less appropriate to predict HABSI. However, when a sign is defined as “an increase of” or “acute onset of,” the clinical relevance is emphasized. Suggestions for risk stratification based on setting, birth weight, or gestational age could also be considered in this context.

### TABLE 3 Pooled ORs and Statistical Heterogeneity of Clinical Parameters Predictive for Health Care–Associated Bloodstream Infections in All Studies

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Pooled OR</th>
<th>95% CI</th>
<th>P</th>
<th>Heterogeneity (I²), %</th>
<th>Studies/Neonates, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical signs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apnea/bradycardia</td>
<td>1.60</td>
<td>0.78–3.30</td>
<td>.20</td>
<td>69.0</td>
<td>5/579</td>
</tr>
<tr>
<td>Distended abdomen</td>
<td>1.30</td>
<td>0.80–2.12</td>
<td>.29</td>
<td>0.0</td>
<td>4/437</td>
</tr>
<tr>
<td>Feeding intolerance</td>
<td>1.40</td>
<td>0.64–3.03</td>
<td>.40</td>
<td>55.0</td>
<td>5/579</td>
</tr>
<tr>
<td>Fever</td>
<td>1.72</td>
<td>0.86–3.45</td>
<td>.13</td>
<td>37.0</td>
<td>3/302</td>
</tr>
<tr>
<td>Grunting</td>
<td>1.02</td>
<td>0.54–1.93</td>
<td>.94</td>
<td>0.0</td>
<td>4/479</td>
</tr>
<tr>
<td>Hypothermia</td>
<td>1.78</td>
<td>0.85–3.74</td>
<td>.13</td>
<td>0.0</td>
<td>3/302</td>
</tr>
<tr>
<td>Lethargy</td>
<td>3.98</td>
<td>1.89–9.40</td>
<td>.002</td>
<td>72.0</td>
<td>4/499</td>
</tr>
<tr>
<td>Pallor/mottling</td>
<td>2.55</td>
<td>1.28–5.18</td>
<td>.010</td>
<td>52.0</td>
<td>3/319</td>
</tr>
<tr>
<td>Tachycardia</td>
<td>1.64</td>
<td>0.86–2.79</td>
<td>.07</td>
<td>5.0</td>
<td>3/399</td>
</tr>
<tr>
<td>Tachypnea</td>
<td>1.27</td>
<td>0.68–2.34</td>
<td>.45</td>
<td>29.0</td>
<td>3/399</td>
</tr>
<tr>
<td><strong>Risk factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPD</td>
<td>5.17</td>
<td>3.35–8.15</td>
<td>.24</td>
<td>79.0</td>
<td>2/180</td>
</tr>
<tr>
<td>Lipid infusion</td>
<td>4.17</td>
<td>1.85–9.40</td>
<td>.001</td>
<td>0.0</td>
<td>2/180</td>
</tr>
<tr>
<td>Male gender</td>
<td>1.27</td>
<td>0.72–2.23</td>
<td>.41</td>
<td>0.0</td>
<td>2/242</td>
</tr>
<tr>
<td>Necrotizing enterocolitis</td>
<td>2.59</td>
<td>0.39–17.50</td>
<td>.33</td>
<td>42.0</td>
<td>2/180</td>
</tr>
<tr>
<td>Postnatal corticosteroids</td>
<td>3.77</td>
<td>1.08–13.18</td>
<td>.04</td>
<td>21.0</td>
<td>2/180</td>
</tr>
<tr>
<td>TPN</td>
<td>3.60</td>
<td>1.69–7.69</td>
<td>.001</td>
<td>0.0</td>
<td>2/180</td>
</tr>
<tr>
<td>Ventilation</td>
<td>1.32</td>
<td>0.43–4.02</td>
<td>.63</td>
<td>70.0</td>
<td>3/322</td>
</tr>
<tr>
<td>Very LBW</td>
<td>2.90</td>
<td>0.38–23.60</td>
<td>.99</td>
<td>87.0</td>
<td>2/180</td>
</tr>
</tbody>
</table>

BPD, bronchopulmonary dysplasia; LBW, low birth weight.

**Random Effects Meta-analysis**

Lethargy and pallor/mottling are significant clinical signs predictive for HABSI (P < .050) and are observed with a high statistical heterogeneity of ≥50%; post hoc analysis with low gestational age neonates revealed a reduction to 0%, thus heterogeneity might be caused by clinical differences. Caution is warranted in interpretation regarding post hoc analysis, but omitting the results of Okascharoen et al. which showed the strongest association for lethargy and pallor/mottling, can also be interpreted as a sensitivity analysis. Therefore, it is likely that our results...
are reliable. Lethargy is often found to be associated with sepsis for all neonates, including low gestational age or very low birth weight neonates.\textsuperscript{6,44,45,54,55} Regarding pallor/mottling, several studies included pallor and/or mottling in their research. One study,\textsuperscript{56} which included all neonates, could not find a significant association between pallor and sepsis. Another multicenter study,\textsuperscript{45} which also included all neonates, clustered pallor and/or motting under impaired peripheral perfusion. In this study, impaired peripheral perfusion was a strong predictor for a positive blood culture (OR 8.5, CI 5.4–13.2). In Ohlin et al’s\textsuperscript{52} study, pallor/mottling was clustered with hypotension, and a significant association was found (OR 2.47, CI 1.48–4.14). Van den

![Table 4](image)

![Figure 3](image)

are reliable. Lethargy is often found to be associated with sepsis for all neonates, including low gestational age or very low birth weight neonates.\textsuperscript{6,44,45,54,55} Regarding pallor/mottling, several studies included pallor and/or motting in their research. One study,\textsuperscript{56} which included all neonates, could not find a significant association between pallor and sepsis. Another multicenter study,\textsuperscript{45} which also included all neonates, clustered pallor and/or motting under impaired peripheral perfusion. In this study, impaired peripheral perfusion was a strong predictor for a positive blood culture (OR 8.5, CI 5.4–13.2). In Ohlin et al’s\textsuperscript{52} study, pallor/mottling was clustered with hypotension, and a significant association was found (OR 2.47, CI 1.48–4.14). Van den
Bruel et al. systematically reviewed clinical signs and their diagnostic value in ambulatory care settings for children aged 1 month to 18 years. They described impaired peripheral perfusion resulting in pallor/mottling and lethargy/reduced consciousness as red flags in the detection of serious infection. Additionally, convulsions, rapid breathing, and cyanosis were strongly associated with serious infection.

Concerning risk factors, TPN, lipid infusion, and postnatal corticosteroid administration were significant after pooling. TPN and lipid infusion are generally accepted as important risk factors, particularly for HABSIs caused by coagulase-negative Staphylococcus.\(^{50,58-62}\) Interestingly, in our analysis, very low birth weight was not a significant predictor despite being associated with HABSI in many studies.\(^{2,61,63}\) The pooled very low birth weight OR involved 2 studies, Okascharoen et al.\(^{31}\) (unadjusted OR 9.1, CI 3.0–26.5) and Mahieu et al.\(^{27}\) (unadjusted OR 1.1, CI 0.5–2.4), representing 180 neonates, 60 of whom had laboratory-confirmed HABSI. The lack of power might be a concern here.

### Limitations and Recommendations

Although a systematic search was done by 2 independent researchers, incomplete retrieval of studies, as well as publication bias and heterogeneity, might influence our results. Concerning the power of the random effects meta-analysis, some pooled parameters are based on <200 suspected cases or <100 laboratory-confirmed cases. For future research, it might be interesting to include all studies in which the objective was to find clinical parameters predictive for HABSI or nosocomial sepsis by univariate analysis. As such, power of the pooled results will increase.

In the past decade, much research has considered heart rate characteristics as an early diagnostic sign for neonatal sepsis.\(^{64-68}\) Including an index of heart rate characteristics in a prediction model for neonatal sepsis seems promising.\(^{64,66,69,70}\)

Differing characteristics of the neonatal populations in various settings are not always reported; this may lead to concerns with external validity of the prediction models. The addition of clinical parameters related to specific neonatal services...
and subgroups is recommended for future work.

Furthermore, future models might potentially benefit from bedside technology capable of quantifying subjective symptoms. For example, pallor/mottling could be replaced by continuous monitoring of peripheral microcirculation, and lethargy could be measured by continuous functional monitoring of brain activity.

CONCLUSIONS

A prediction model should be considered as guidance rather than an absolute indicator because all models have limited diagnostic accuracy. However, the use of prediction models might decrease the use of antimicrobial therapy. In clinical practice, a NOSEP score of 8 or more by Mahieu et al27 has the greatest potential but may necessitate additional consideration of the patient’s clinical condition, depending on the criteria for culturing in a given setting. Furthermore, these findings suggest that lethargy and pallor/mottling for all neonates and apnea and/or bradycardia and poor peripheral perfusion for very low birth weight neonates are the most powerful clinical signs in the prediction of HABSI. Nonetheless, other clinical signs should not be discarded. As stated earlier, the particular clinical context should always be taken into account.

ACKNOWLEDGMENT

The authors thank Dr Ellen Deschepper for her biostatistical advice, particularly regarding diagnostic accuracy.

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Evelien Verstraete is supported by a Special Research Fund at Ghent University. Stijn Blot holds a research mandate of the Special Research Fund at Ghent University.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

16. Chiesa C, Panero A, Osborn JF, Simonetti AF, Pacifico L. Diagnosis of neonatal


Prediction Models for Neonatal Health Care–Associated Sepsis: A Meta-analysis
Evelien Hilde Verstraete, Koen Blot, Ludo Mahieu, Dirk Vogelaers and Stijn Blot
Pediatrics 2015;135;e1002; originally published online March 9, 2015;
DOI: 10.1542/peds.2014-3226

The online version of this article, along with updated information and services, is
located on the World Wide Web at:
/content/135/4/e1002.full.html