

## REVIEW ARTICLE

# Artificial and human intelligence for early identification of neonatal sepsis

 Brynne A. Sullivan<sup>1✉</sup>, Sherry L. Kausch<sup>1</sup> and Karen D. Fairchild<sup>1</sup>

© The Author(s), under exclusive licence to the International Pediatric Research Foundation, Inc 2022

Artificial intelligence may have a role in the early detection of sepsis in neonates. Machine learning can identify patterns that predict high or increasing risk for clinical deterioration from a sepsis-like illness. In developing this potential addition to NICU care, careful consideration should be given to the data and methods used to develop, validate, and evaluate prediction models. When an AI system alerts clinicians to a change in a patient's condition that warrants a bedside evaluation, human intelligence and experience come into play to determine an appropriate course of action: evaluate and treat or wait and watch closely. With intelligently developed, validated, and implemented AI sepsis systems, both clinicians and patients stand to benefit.

*Pediatric Research* (2023) 93:350–356; <https://doi.org/10.1038/s41390-022-02274-7>

**IMPACT:**

- This narrative review highlights the application of AI in neonatal sepsis prediction. It describes issues in clinical prediction model development specific to this population.
- This article reviews the methods, considerations, and literature on neonatal sepsis model development and validation.
- Challenges of AI technology and potential barriers to using sepsis AI systems in the NICU are discussed.

Sepsis continues to cause significant morbidity and mortality among preterm very low birthweight (VLBW) infants in the neonatal intensive care unit (NICU), and earlier detection and treatment can reduce mortality and improve outcomes for survivors. In this narrative review, we address a number of questions related to artificial intelligence (AI) for sepsis prediction and detection in NICU patients. First, we discuss aspects of neonatal sepsis that make it a tractable problem for machine learning (ML) predictive models. Next, we cover technical aspects of ML model development and validation, including variable selection using both static and dynamic data. We then review some existing early warning and ML systems. Finally, we discuss the benefits of and barriers to implementing sepsis prediction systems in the NICU, with the goal of “right timing” antibiotics to improve patient outcomes.

## Q1: HOW SUITABLE IS THE PROBLEM OF NEONATAL SEPSIS FOR AI SOLUTIONS?

Premature infants in the NICU are, in a number of ways, an ideal population for AI-based sepsis monitoring. They are immune-compromised and require invasive devices that create a high risk for sepsis, yet they may have a period of relative stability before developing sepsis. Late-onset sepsis (LOS) does not present when the pathogen invades the blood stream, but instead as a sub-acute physiologic response with inflammation and organ dysfunction. Therefore, when advanced analytics of patient-generated data can detect the transition from “well” to “ill”, predictive models can translate this information to the clinical team to provide

earlier warning of a sub-acute, potentially catastrophic deterioration. The advantage of early warning and treatment of sepsis must be considered in balance with the potential disadvantage of increasing antibiotic exposure, which has negative consequences.<sup>1–5</sup> Non-specific signs of sepsis such as apnea and respiratory distress and the risk of rapid deterioration with delayed treatment make this balance challenging for clinicians. Thus, it is imperative to not only develop sepsis warning systems with limited false alarms, but also to teach clinicians to use AI model output in the context of all available clinical data in making decisions about starting and stopping antibiotics. A final consideration is the distinction between early- and late-onset sepsis (EOS within 3 days from birth, LOS after 3 days). For EOS, a simple static prediction model (the EOS calculator) has been developed and its broad implementation has reduced antibiotic use.<sup>6</sup> In this review we focus primarily on the prediction of LOS incorporating both static and dynamic data, including continuously streaming vital sign data from NICU bedside monitors.<sup>7</sup>

Prediction models will perform best if the targeted outcome is well-defined and validated. For sepsis, this requires careful medical record review rather than simply relying on ICD codes since many studies have shown that diagnostic codes for sepsis are inaccurate.<sup>8,9</sup> A challenge with regard to neonatal sepsis is the lack of a consensus definition,<sup>10,11</sup> making it difficult to compare and interpret results across studies.<sup>12,13</sup> Some prediction models train only on culture-positive sepsis, while many include cases of “clinical sepsis” in which an infant has significant signs of illness and clinicians opt to prolong antibiotic treatment despite negative cultures. Experts argue that in the setting of modern laboratory

<sup>1</sup>Department of Pediatrics, University of Virginia School of Medicine, Charlottesville, VA, USA. ✉email: bsa4m@virginia.edu

Received: 18 April 2022 Revised: 29 July 2022 Accepted: 5 August 2022

Published online: 20 September 2022

equipment and sufficient inoculation volume, the likelihood of a false negative blood culture is extremely low.<sup>14,15</sup> Nonetheless, as discussed later in this review, many prediction models have been developed using both clinical and culture-positive sepsis cases, and it is therefore important for clinicians to use judgment to decide on the duration of therapy in the face of a high or rising risk score, since misuse of antibiotics can lead to adverse outcomes.<sup>16</sup>

Finally, for AI models to be widely useful for NICU patients they must be generalizable and reproducible. FAIR data principles were proposed as a way for AI research and development to achieve this goal—data should be findable, accessible, interoperable, and reusable.<sup>17,18</sup> Large data sets and external validation are likely to improve generalizability and translation to clinical care. However, generating data that are FAIR and models that are externally validated in large cohorts is no small task; it typically requires long-standing multicenter, multi-specialty research collaborations.

**Q2: WHAT ARE THE IMPORTANT AI AND MACHINE LEARNING MODEL DEVELOPMENT CONCEPTS?**

Figures 1 and 2 provide a conceptual overview of aspects of AI relevant to healthcare applications, from algorithm development through clinical implementation and integration. ML is a type of AI that includes supervised methods such as classification and regression, using algorithms to find structure in labeled data, and unsupervised methods involving clustering and dimension reduction of unlabeled data. Generally, sepsis prediction models use supervised ML with various modeling methods, including regression, tree-based methods, neural networks, and others. In some studies, a variety of modeling methods were shown to have similar predictive performance,<sup>19,20</sup> while in other studies, a specific method is found to have better performance.<sup>21</sup>

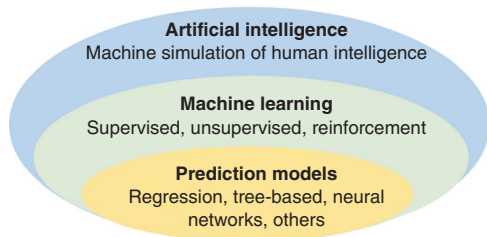
A common way to assess ML model performance is using the area under the receiver operating characteristics curve (AUC) to summarize the model’s ability to discriminate cases from controls over all possible thresholds.<sup>22</sup> The AUC value alone is insufficient

to evaluate model performance since it does not consider prior probability, does not provide information about the distribution of errors, and weights omission and commission errors equally.<sup>23,24</sup> Moreover, even a model with good discrimination may provide risk estimates that are unreliable.<sup>25</sup>

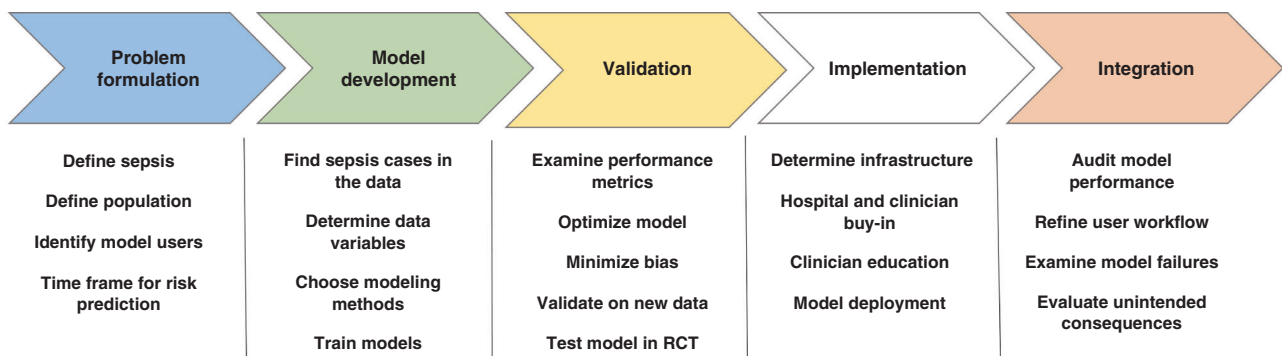
Another way to evaluate model performance is by calculating sensitivity, specificity, and negative and positive predictive values (NPV and PPV).<sup>26</sup> Although LOS occurs in approximately 15% of very preterm infants, the chance of an infant developing sepsis on any particular day is quite low. Thus, the PPV of models developed to continuously evaluate the risk of imminent sepsis will be low in order to have acceptably high sensitivity. ML model performance should also be evaluated using qualitative methods, such as calibration plots and time-to-event plots (Fig. 3). The calibration of a model’s risk predictions can be visualized by plotting the observed risk as a function of the predicted risk.<sup>27</sup> Time-to-event plots show the average model output in a cohort relative to the time of the event and illustrate the horizon or lead time for sepsis prediction. This qualitative model assessment provides valuable information about its clinical utility since a score without a rise before clinicians recognize illness is not likely to benefit patients.

Once a model is developed or trained, testing or validation is an essential next step. A validation data set can be internal (a subset of the original data set) or external, from a new cohort at a different center. Validation in data sets with similar patient characteristics and practice patterns compared to the training data provides evidence for reproducibility of model performance, while external validation using data from cohorts with different characteristics (for example, different centers, patient demographics, level of illness, or clinical practices), provides evidence of model transportability.<sup>28</sup> In an example from our prior work, we showed that differences in invasive versus non-invasive respiratory support across NICUs impacted the performance of a sepsis prediction model that incorporated features to detect apnea.<sup>29</sup> In addition to external validation, ongoing evaluation of ML models ensures adequate performance after implementation. Data shift or drift may occur over time as practices, hospital systems, and patient populations change.<sup>30</sup> Examples that could impact NICU sepsis model performance include a change in bedside monitors with differences in HR or SpO<sub>2</sub> averaging times, change in practices for obtaining specific laboratory tests that serve as model inputs, or changes in the use of medications or respiratory support that may impact vital sign patterns.

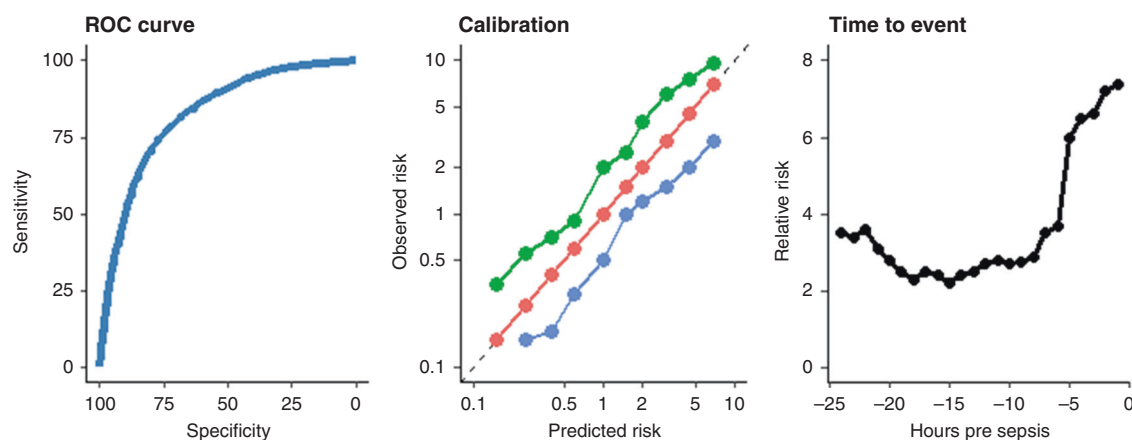
The ultimate step in model evaluation is conducting randomized clinical trials to determine whether displaying the output of an AI algorithm leads to meaningfully improved outcomes. Only through well-designed, large clinical trials will sepsis AI systems be trusted, implemented, and routinely used for patient care. Finally, since many research groups are developing sepsis AI, it is important that results and algorithms



**Fig. 1 Concepts overview.** Schema illustrating the overlapping concepts of artificial intelligence, machine learning, and prediction models with brief descriptions.



**Fig. 2 Artificial intelligence (AI) process diagram.** A conceptual diagram illustrating the key steps of developing sepsis AI technology, from idea to model development, testing, and translation to clinical implementation.



**Fig. 3 Examples of model performance metrics.** The left panel displays an area under the receiver operating characteristic curve for a population of patients. The AUC is created by plotting the sensitivity against the specificity across all thresholds of the model output. The middle panel is an example of three model calibration curves for a population of patients. Predicted risk relative to average is on the abscissa and observed risk relative to average is on the ordinate. Each point represents one decile of predicted risk. The line of identity is shown as a dashed line and represents a well-calibrated model. The other two lines represent models that either over- or under-predict risk. The right panel is an example of a “time-to-event” plot for an individual patient. In this example, the model output (risk of sepsis score) rises steeply 4 h before a patient had a blood culture sent that diagnosed sepsis at time zero. Theoretically, if clinicians could see and interpret the rising score, antibiotics might have been given 4 h earlier in this example.

be shared among researchers. In order to better interpret results across studies, models should be reported using a standardized format such as TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis)<sup>31</sup> or a subsequent format specific for AI (TRIPOD-AI).<sup>32</sup>

### Q3: WHICH DATA ARE USEFUL IN NEONATAL SEPSIS PREDICTION MODELS?

When predicting imminent LOS, AI models may use physiologic data derived from high-resolution data (e.g., the electrocardiogram waveform signal sampled at 250 Hz), low-resolution data (e.g., demographics, clinical risks or signs, intermittently sampled vital signs, or laboratory tests), or a combination of both. The inflammatory response to sepsis manifests as changes in multiple physiologic processes that we measure as vital signs, making these data particularly useful for detecting and predicting LOS.<sup>33,34</sup> Patterns in continuous cardiorespiratory data have been identified as signatures of illness due to sepsis. Predictive modeling translates these physiologic data into a predicted risk of imminent deterioration, which has potential clinical utility as an early warning system. For example, low variability of HR accompanied by HR decelerations was recognized as a signature of illness in neonates.<sup>35</sup> The mechanism of abnormal heart rate characteristics (HRC) during sepsis involves cytokine signaling and autonomic nervous system activation with increased vagus nerve firing.<sup>36–39</sup> The HRC index, a continuous sepsis prediction model, was developed to capture these abnormal patterns and is discussed later in this review.<sup>7</sup>

In searching beyond HR patterns for physiologic markers of sepsis, a logical place to look is in the respiratory data. An increase in central apnea is one of the major signs of sepsis in preterm NICU patients,<sup>38–42</sup> due in part to the cytokine-triggered release of endogenous prostaglandins.<sup>38–40</sup> Apnea detection through chest impedance waveform analysis is complicated, while detection of a decline in HR and SpO<sub>2</sub> that often accompany apnea is simpler. One analytic that serves this purpose is the cross-correlation of HR and SpO<sub>2</sub>, which measures the degree to which the two signals co-trend within a set lag time. An increase in this metric captures deceleration-desaturation events, which correlate with increased central apnea or exaggerated pathologic periodic breathing in preterm infants.<sup>29,41</sup>

Changes in physiologic data can be non-specific while still useful for sepsis detection and prediction. Interpreting a rising sepsis risk score may require consideration of the clinical context and the patient’s baseline condition. Some preterm infants have chronically abnormal HR and SpO<sub>2</sub> patterns reflecting pathologies unrelated to sepsis. One solution could be to incorporate a patient’s baseline into the calculated risk to account for inter- and intra- patient variability and allow for personalized AI predictions.

### Demographic, laboratory, and clinical data for sepsis AI algorithms

The EHR contains many pertinent clinical data that add to sepsis risk prediction. Lower gestational age and birthweight correlate strongly with rising risk of LOS, and can stand alone to risk stratify premature infants at birth or add to models that use continuous vital sign data.<sup>7,43–45</sup> Postnatal and postmenstrual age also add to risk prediction due to the peak in LOS incidence at 1–3 weeks of age.<sup>46,47</sup> Additional demographic and perinatal variables may improve model performance, such as sex,<sup>48,49</sup> race, ethnicity, or delivery mode.<sup>50,51</sup> While including these variables in ML model is likely to improve the AUC, they provide only static information.

Laboratory tests that measure components of the host response to infection, such as immature neutrophils or serial C-reactive protein values are commonly used tests for sepsis screening and may serve as decision support for either starting or withholding antibiotics in conjunction with other clinical variables.<sup>52–56</sup> However, such tests are typically ordered by the clinician with concern for sepsis and therefore the information they provide is likely to lag behind clinical suspicion rather than provide early warning.

Clinical risk factors may be incorporated into sepsis ML models including the presence of a central vascular catheter or mechanical ventilation and medications that increase sepsis risk such as postnatal steroids. Clinical signs of sepsis may also be incorporated into models, including increased apnea, respiratory distress, feeding intolerance, poor perfusion, temperature instability, hypotension, and lethargy.<sup>42,57,58</sup> These signs may be captured from EHR documentation, but once an ICU clinician documents “lethargy” in the EHR they typically have already ordered the blood culture and antibiotics. Several published sepsis detection models instead use patient-generated data to detect clinical signs of sepsis, including using HR and SpO<sub>2</sub> data to detect an increase

in apnea,<sup>41</sup> core to peripheral temperature differential to detect impaired thermoregulation,<sup>59</sup> and cardiorespiratory waveform data to detect decreased infant motion or lethargy.<sup>60</sup> AI models for LOS in the NICU that include continuous physiologic data are likely to be more clinically useful than those that use only EHR data.

#### Q4: WHAT ADVANCED WARNING SYSTEMS FOR SEPSIS EXIST, AND WHAT LIES IN THE FUTURE?

Before discussing AI systems for sepsis, consideration should be given to “Early Warning Scores” (EWS). EWS and ML models are both designed to alert the medical team to concerning clinical changes that might otherwise go unnoticed. Both can be integrated into the EHR or displayed at the bedside, and both can incorporate information from a mix of static and dynamic clinical variables. EWS employ a “track and trigger” approach, whereas AI models use math and the data to learn temporal trends and correlation among parameters.<sup>61</sup> For example, the Pediatric Early Warning Score is calculated based on periodic observations of multiple physiological parameters and designed to predict clinical deterioration (including but not limited to sepsis) in hospitalized children.<sup>62,63</sup> AI models would be expected to perform better than EWS because they can use the data as continuous rather than categorical values determined by thresholds. Additionally, modeling rather than empirically derived cutoffs can detect more subtle and complex patterns in the data associated with the target outcome.

Though this review is focused on neonatal LOS prediction, the EOS calculator deserves mention since it is widely used and exemplifies some important aspects of sepsis AI.<sup>6,64,65</sup> The model uses perinatal risk factors known at the time of birth in a logistic regression model to derive prior probability and then incorporates the clinicians’ assessment (asymptomatic, equivocal, or clinically ill) using Bayes’ theorem. The risk per 1000 live births is displayed for each of the three categories of illness.<sup>6,64</sup> Decision support is provided, allowing for clinical judgment to guide the application of the AI technology, which is likely a factor in the widespread adoption of this model. Studies of the impact of the EOS calculator have shown it reduces the number of asymptomatic or equivocal infants with sepsis risk factors undergoing laboratory evaluations and exposure to antibiotics.<sup>64</sup>

Developing a calculator for LOS would be substantially more complicated, since it is more common than EOS, occurs over a wide time range, and has non-specific clinical signs that are common in preterm infants with non-infectious conditions. Nonetheless, a number of tools for predicting LOS in NICU patients before they are obviously sick have been published.<sup>66</sup> Table 1 summarizes models using continuous or intermittently sampled data to predict LOS before clinical deterioration prompting a blood culture and antibiotics.<sup>7,60,67–69</sup> Several other studies have used data at the time of blood culture to predict whether sepsis will be ruled in or out (positive versus negative culture) which may be useful for determining when to start and stop antibiotics.<sup>70,71</sup> And finally, several studies have used vital sign data shortly after birth to predict the risk of developing sepsis later in the NICU course, which might identify highest risk infants in need of enhanced vigilance or preventive strategies not suitable for the entire preterm population.<sup>72,73</sup>

To date, the only commercially available system for NICU predictive monitoring using continuous bedside monitor data is the HRC index, or HeRO Score. This is also the only ML model for neonatal sepsis that has been tested in a randomized clinical trial and shown to improve important clinical outcomes.<sup>74</sup> The HRC algorithm uses electrocardiogram data from standard NICU bedside monitors to calculate the fold-increased risk of a clinical deterioration due to sepsis (culture-proven) or a sepsis-like illness (clinical sepsis) in the next 24 h. The algorithm uses mathematical

**Table 1.** A summary of select studies reporting the development and performance of machine learning (ML) models to predict imminent late-onset sepsis.

First author (model name)	Population	Late-onset sepsis cases used for modeling	Data input	Sepsis training window (h)	External validation	Modeling method
Griffin et al. (HRC index) <sup>7,74,89</sup>	VLBW	Culture positive and clinical	ECG heart rate characteristics (low variability, decels)	–24 to 0	Yes	Logistic regression
Gur et al. (RALIS) <sup>67,90,91</sup>	VLBW	Culture positive and clinical	HR, RR, temp, weight, desaturation, bradycardia episodes	Not reported	Not reported	Not specified
Masino et al. <sup>93</sup>	All NICU patients	Culture positive and clinical	36 features from EHR, including VS and clinical status	–48 to –4	No	8 ML models
Mani et al. <sup>69</sup>	All NICU patients	Culture positive and clinical	EHR data: 71 laboratory 30 demographic + clinical variables	–48 to –12	No	9 ML models
Joshi, et al. <sup>60</sup> Cabrera-Quiros, et al. <sup>92</sup>	<32 weeks GA	Culture positive	4 feature subsets: (1) HRV, (2) movement, (3) respiration, and (4) combination of above	–24 to 0 and –3 to 0	No	Logistic regression, naive Bayes, nearest mean classifier
Song et al. <sup>68</sup>	All NICU patients	Culture positive and clinical	Vital signs, blood gas data, CBC	–24 to 0	No	8 ML models

calculations that report on decreased HR variability and transient HR decelerations, patterns shown in pre-clinical models to reflect pathogen-induced inflammatory cytokine release and vagus nerve firing.<sup>37,75</sup> In a randomized clinical trial of 3003 VLBW infants at nine NICUs,<sup>74</sup> display of this risk score was associated with significantly lower sepsis-associated mortality (12% versus 20%), presumably due to earlier treatment.<sup>76</sup> Importantly, the display of the score resulted in a small increase in the number of blood cultures and antibiotic days, but only among infants with confirmed sepsis.<sup>76</sup> This indicated that clinicians may have also used the score for its NPV to decide not to start antibiotics or to discontinue antibiotics in patients with non-specific, mild clinical signs.

#### Q5: WHAT ARE SOME BENEFITS AND BARRIERS TO SEPSIS ML MODEL IMPLEMENTATION AND CLINICAL INTEGRATION?

Much has been written about the potential benefits of AI implementation in healthcare<sup>77</sup> but “AI solutions” will not replace the hard work of clinicians deciding which patients require testing and therapies.<sup>78</sup> Properly developed sepsis AI systems might direct clinicians to the right bed at the right time, leading to earlier antibiotic treatment and supportive care leading to improved outcomes. The 20% reduction in sepsis-associated mortality with continuous HRC index display in the HeRO RCT is an example. For survivors of neonatal sepsis, earlier treatment might have other benefits, such as reduced NICU length of stay.<sup>79</sup> The caveat, of course, is that attention must be given when implementing sepsis AI to avoid misuse of antibiotics for non-infectious clinical deterioration that is common in preterm infants in the NICU.

Beyond direct patient benefits, other potential benefits of using AI risk models for sepsis include resource allocation and risk stratification, which can be useful for cost-effectiveness analyses, classification for research, and benchmarking across hospitals. Also, care is required in AI model design to avoid introducing biased data into algorithms. A first step in addressing bias in AI is to develop and test algorithms for performance across the spectrum of patient sex, race, ethnicity, and socioeconomic status. With regard to sepsis, a potential advantage of physiology-based algorithms is that heart rate patterns of neonates tend to be similar across the spectrum of patient diversity. Adding pulse oximetry data to HRC will need close scrutiny since racial differences in accuracy of pulse oximetry data have recently been described in adults,<sup>80</sup> children,<sup>81</sup> and neonates.<sup>82</sup> Regardless of what sources of data serve as model input, AI algorithms should be developed and validated in large, diverse patient populations with efforts made to minimize bias of all types.

Although there are many potential benefits of AI, there are also many barriers. We developed the acronym “BARRIERS” to summarize some major challenges in this field: Babies, Analytics, Reactors, Reassurance, Integration, Equipment, Re-education, and Space.<sup>83</sup> Babies themselves can complicate the development and deployment of early warning systems for sepsis since they cannot announce that they feel sick, and their signs of sepsis are non-specific and overlap with normal preterm physiology. This creates the problem of false alarms in a unit already prone to alarm fatigue.<sup>84,85</sup> Another barrier, “Analytics,” refers to the difficulty in creating models due to heterogeneity of event identification, variable selection, and modeling techniques, as previously discussed in this review. “Reactors” are the model users, NICU clinicians with varied education, experience, and responsibilities. The barrier, in this case, is the difficulty in displaying data and clinical decision support in a way that is effective for a broad range of clinicians. “Reassurance” can be a problem with AI models if a low-risk score falsely reassures the clinical team faced with an infant with significant signs of illness, leading to a delay in treatment. “Integration” refers to the challenge of introducing an AI model without creating too many distracting false alarms. One

way to mitigate alarm fatigue yet assure that critical information is transmitted to the right person is to have a centralized clinical team that reviews alerts and determines which ones should be transmitted to the care team,<sup>86</sup> an approach that may not be broadly feasible. The “E” in BARRIERS is equipment that must be integrated into the clinical workflow. Once the system is integrated, education and “Re-education” for users are critically important to assure proper implementation. And finally, “Space” can be a barrier since the NICU bedside may already be crowded with equipment and monitors. A new sepsis prediction system needs to be positioned in such a way to be noticeable but not overwhelming.

In the case of an AI system that is shown to improve patient outcomes, implementation relies on hospital administrators and clinicians “buying in.” A survey-based study of continuous predictive monitoring reported that users had positive engagement with the system if they trusted the data used in the model and if they understood the science behind the model outputs.<sup>87</sup> This is the basis for the term “explainable AI” which some view as essential for clinicians to utilize the system, although others argue that methods to make models explainable sacrifice their performance.<sup>88</sup> A final consideration is that AI systems may introduce unintended consequences such as inappropriate testing and therapies. Further research is needed to characterize (or differentiate) how sepsis AI models perform in events of non-specific clinical deterioration versus culture-proven sepsis.

#### CONCLUSION

Sepsis AI is a way to analyze and present data to clinicians for earlier detection and treatment leading to improved patient outcomes. If properly developed and implemented, AI systems can alert clinicians to a change in a patient’s condition that warrants a bedside evaluation. At that point, human intelligence and experience can combine computer-generated risk information with what they see and what they know to make the best decisions for individual patients.

#### REFERENCES

1. Gasparrini, A. J. et al. Antibiotic perturbation of the preterm infant gut microbiome and resistome. *Gut Microbes* **7**, 443–449 (2016).
2. Dardas, M. et al. The impact of postnatal antibiotics on the preterm intestinal microbiome. *Pediatr. Res.* **76**, 150–158 (2014).
3. Pammi, M. et al. Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis. *Microbiome* **5**, 31 (2017).
4. Cantey, J. B., Pyle, A. K., Wozniak, P. S., Hynan, L. S. & Sánchez, P. J. Early antibiotic exposure and adverse outcomes in preterm, very low birth weight infants. *J. Pediatr.* **203**, 62–67 (2018).
5. Ting, J. Y. et al. Association between antibiotic use and neonatal mortality and morbidities in very low-birth-weight infants without culture-proven sepsis or necrotizing enterocolitis. *JAMA Pediatr.* **170**, 1181–1187 (2016).
6. Escobar, G. J. et al. Stratification of risk of early-onset sepsis in newborns  $\geq 34$  weeks’ gestation. *Pediatrics* **133**, 30–36 (2014).
7. Griffin, M. P. et al. Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness. *Pediatr. Res.* **53**, 920–926 (2003).
8. Iwashyna, T. J. et al. Identifying patients with severe sepsis using administrative claims: patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. *Med. Care* **52**, e39–e43 (2014).
9. Ramanathan, R. et al. Validity of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) screening for sepsis in surgical mortalities. *Surg. Infect. (Larchmt.)* **15**, 513–516 (2014).
10. Wynn, J. L. & Polin, R. A. Progress in the management of neonatal sepsis: the importance of a consensus definition. *Pediatr. Res.* **83**, 13–15 (2018).
11. Molloy, E. J. et al. Neonatal sepsis: need for consensus definition, collaboration and core outcomes. *Pediatr. Res.* **88**, 2–4 (2020).
12. Henry, C. J. et al. Neonatal sepsis: a systematic review of core outcomes from randomised clinical trials. *Pediatr. Res.* **91**, 735–742. <https://doi.org/10.1038/s41390-021-01883-y> (2022).
13. Hayes, R. et al. Neonatal sepsis definitions from randomised clinical trials. *Pediatr. Res.* <https://doi.org/10.1038/s41390-021-01749-3> (2021).

14. Cantey, J. B. & Baird, S. D. Ending the culture of culture-negative sepsis in the neonatal ICU. *Pediatrics* **140**, e20170044 (2017).
15. Cantey, J. B. & Prusakov, P. A proposed framework for the clinical management of neonatal "culture-negative" sepsis. *J. Pediatr.* **244**, 203–211. <https://doi.org/10.1016/j.jpeds.2022.01.006> (2022).
16. Mukhopadhyay, S. & Puopolo, K. M. Antibiotic use and mortality among premature infants without confirmed infection-perpetrator or innocent bystander? *JAMA Pediatr.* **170**, 1144–1146 (2016).
17. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
18. Levinson, M. A. et al. FAIRSCAPE: a framework for FAIR and reproducible biomedical analytics. *Neuroinformatics* <https://doi.org/10.1007/s12021-021-09529-4> (2021).
19. Beaulieu-Jones, B. K. et al. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? *npj Digital Med.* **4**, 62 (2021).
20. Beam, A. L. & Kohane, I. S. Translating artificial intelligence into clinical care. *JAMA* **316**, 2368–2369 (2016).
21. Spaeder, M. C. et al. Perioperative near-infrared spectroscopy monitoring in neonates with congenital heart disease: relationship of cerebral tissue oxygenation index variability with neurodevelopmental outcome. *Pediatr. Crit. Care Med.* **18**, 213–218 (2017).
22. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* 106–119 (Springer New York, 2009). <https://doi.org/10.1007/978-0-387-84858-7>
23. Justice, A. C., Covinsky, K. E. & Berlin, J. A. Assessing the generalizability of prognostic information. *Ann. Intern. Med.* **130**, 515–524 (1999).
24. Harrell, F. E., Lee, K. L. & Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* **15**, 361–387 (1996).
25. Diamond, G. A. What price perfection? Calibration and discrimination of clinical prediction models. *J. Clin. Epidemiol.* **45**, 85–89 (1992).
26. Pinker, E. Reporting accuracy of rare event classifiers. *npj Digital Med.* **1**, 56 (2018).
27. Harrell, F. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis (Springer Series in Statistics)* (Springer International Publishing, 2015).
28. Nevin, L., PLOS Medicine Editors. Advancing the beneficial use of machine learning in health care and medicine: toward a community understanding. *PLoS Med.* **15**, e1002708 (2018).
29. Fairchild, K. D. et al. Vital signs and their cross-correlation in sepsis and NEC: a study of 1,065 very-low-birth-weight infants in two NICUs. *Pediatr. Res.* **81**, 315–321 (2017).
30. Subbaswamy, A., Adams, R. & Saria, S. Evaluating model robustness and stability to dataset shift. Preprint at *arXiv*. <https://doi.org/10.48550/arxiv.2010.15100> (2020).
31. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann. Intern. Med.* **162**, 55–63 (2015).
32. Collins, G. S. et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* **11**, e048008 (2021).
33. Sullivan, B. A. & Fairchild, K. D. Vital signs as physiologic markers of neonatal sepsis. *Pediatr. Res.* **91**, 273–282. <https://doi.org/10.1038/s41390-021-01709-x> (2022).
34. Kumar, N., Akangire, G., Sullivan, B., Fairchild, K. & Sampath, V. Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront. *Pediatr. Res.* **87**, 210–220 (2020).
35. Sullivan, B. A. & Fairchild, K. D. Predictive monitoring for sepsis and necrotizing enterocolitis to prevent shock. *Semin. Fetal Neonatal Med.* **20**, 255–261 (2015).
36. Tracey, K. J. Physiology and immunology of the cholinergic antiinflammatory pathway. *J. Clin. Invest.* **117**, 289–296 (2007).
37. Fairchild, K. D., Srinivasan, V., Moorman, J. R., Gaykema, R. P. A. & Goehler, L. E. Pathogen-induced heart rate changes associated with cholinergic nervous system activation. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **300**, R330–R339 (2011).
38. Fairchild, K. et al. Clinical associations of immature breathing in preterm infants: part 1—central apnea. *Pediatr. Res.* **80**, 21–27 (2016).
39. Siljehav, V., Hofstetter, A. M., Leifsdottir, K. & Herlenius, E. Prostaglandin E2 mediates cardiorespiratory disturbances during infection in neonates. *J. Pediatr.* **167**, 1207–1213.e3 (2015).
40. Herlenius, E. An inflammatory pathway to apnea and autonomic dysregulation. *Respir. Physiol. Neurobiol.* **178**, 449–457 (2011).
41. Fairchild, K. D. & Lake, D. E. Cross-correlation of heart rate and oxygen saturation in very low birthweight infants: association with apnea and adverse events. *Am. J. Perinatol.* **35**, 463–469 (2018).
42. Das, A., Shukla, S., Rahman, N., Gunzler, D. & Abughali, N. Clinical indicators of late-onset sepsis workup in very low-birth-weight infants in the neonatal intensive care unit. *Am. J. Perinatol.* **33**, 856–860 (2016).
43. Shane, A. L. & Stoll, B. J. Neonatal sepsis: progress towards improved outcomes. *J. Infect.* **68**(Suppl 1), S24–S32 (2014).
44. Shane, A. L., Sánchez, P. J. & Stoll, B. J. Neonatal sepsis. *Lancet* **390**, 1770–1780 (2017).
45. Köstlin-Gille, N. et al. Epidemiology of early and late onset neonatal sepsis in very low birthweight infants: data from the German neonatal network. *Pediatr. Infect. Dis. J.* **40**, 255–259 (2021).
46. Stoll, B. J. et al. Late-onset sepsis in very low birth weight neonates: the experience of the NICHD Neonatal Research Network. *Pediatrics* **110**, 285–291 (2002).
47. Hornik, C. P. et al. Early and late onset sepsis in very-low-birth-weight infants from a large group of neonatal intensive care units. *Early Hum. Dev.* **88**(Suppl 2), S69–S74 (2012).
48. Stevenson, D. K. et al. Sex differences in outcomes of very low birthweight infants: the newborn male disadvantage. *Arch. Dis. Child. Fetal Neonatal Ed.* **83**, F182–F185 (2000).
49. O'Driscoll, D. N., McGovern, M., Greene, C. M. & Molloy, E. J. Gender disparities in preterm neonatal outcomes. *Acta Paediatr.* <https://doi.org/10.1111/apa.14390> (2018).
50. Travers, C. P. et al. Racial/ethnic disparities among extremely preterm infants in the United States from 2002 to 2016. *JAMA Netw. Open* **3**, e206757 (2020).
51. Wallace, M. E. et al. Racial/ethnic differences in preterm perinatal outcomes. *Am. J. Obstet. Gynecol.* **216**, 306.e1–306.e12 (2017).
52. Ohlin, A., Björkqvist, M., Montgomery, S. M. & Schollin, J. Clinical signs and CRP values associated with blood culture results in neonates evaluated for suspected sepsis. *Acta Paediatr.* **99**, 1635–1640 (2010).
53. Coggins, S. A. et al. Use of a computerized C-reactive protein (CRP) based sepsis evaluation in very low birth weight (VLBW) infants: a five-year experience. *PLoS One* **8**, e78602 (2013).
54. Rønnestad, A., Abrahamsen, T. G., Gaustad, P. & Finne, P. H. C-reactive protein (CRP) response patterns in neonatal septicemia. *APMIS* **107**, 593–600 (1999).
55. Benitz, W. E., Han, M. Y., Madan, A. & Ramachandra, P. Serial serum C-reactive protein levels in the diagnosis of neonatal infection. *Pediatrics* **102**, E41 (1998).
56. Brown, J. V. E., Meader, N., Wright, K., Cleminson, J. & McGuire, W. Assessment of C-reactive protein diagnostic test accuracy for late-onset infection in newborn infants: a systematic review and meta-analysis. *JAMA Pediatr.* **174**, 260–268 (2020).
57. Sullivan, B. A. et al. Clinical and vital sign changes associated with late-onset sepsis in very low birth weight infants at 3 NICUs. *J. Neonatal Perinat. Med.* **14**, 553–561 (2021).
58. Fanaroff, A. A. et al. Incidence, presenting features, risk factors and significance of late onset septicemia in very low birth weight infants. The National Institute of Child Health and Human Development Neonatal Research Network. *Pediatr. Infect. Dis. J.* **17**, 593–598 (1998).
59. Knobel-Dail, R. B., Sloane, R., Holditch-Davis, D. & Tanaka, D. T. Negative temperature differential in preterm infants less than 29 weeks gestational age: associations with infection and maternal smoking. *Nurs. Res.* **66**, 442–453 (2017).
60. Joshi, R. et al. Predicting neonatal sepsis using features of heart rate variability, respiratory characteristics, and ECG-derived estimates of infant motion. *IEEE J. Biomed. Health Inf.* **24**, 681–692 (2020).
61. Gao, H. et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med.* **33**, 667–679 (2007).
62. Duncan, H., Hutchison, J. & Parshuram, C. S. The Pediatric Early Warning System score: a severity of illness score to predict urgent medical need in hospitalized children. *J. Crit. Care* **21**, 271–278 (2006).
63. Lambert, V., Matthews, A., MacDonell, R. & Fitzsimons, J. Paediatric early warning systems for detecting and responding to clinical deterioration in children: a systematic review. *BMJ Open* **7**, e014497 (2017).
64. Kuzniewicz, M. W. et al. A quantitative, risk-based approach to the management of neonatal early-onset sepsis. *JAMA Pediatr.* **171**, 365–371 (2017).
65. Puopolo, K. M. et al. Estimating the probability of neonatal early-onset infection on the basis of maternal risk factors. *Pediatrics* **128**, e1155–e1163 (2011).
66. Persad, E. et al. Neonatal sepsis prediction through clinical decision support algorithms: a systematic review. *Acta Paediatr.* **110**, 3201–3226 (2021).
67. Gur, I. et al. Pilot study of a new mathematical algorithm for early detection of late-onset sepsis in very low-birth-weight infants. *Am. J. Perinatol.* **32**, 321–330 (2015).
68. Song, W. et al. A predictive model based on machine learning for the early detection of late-onset neonatal sepsis: development and observational study. *JMIR Med. Inform.* **8**, e15965 (2020).
69. Mani, S. et al. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. *J. Am. Med. Assoc.* **21**, 326–336 (2014).
70. Goldberg, O. et al. Can we improve early identification of neonatal late-onset sepsis? A validated prediction model. *J. Perinatol.* **40**, 1315–1322 (2020).

71. Sweeney, T. E. et al. Validation of the sepsis metascore for diagnosis of neonatal sepsis. *J. Pediatr. Infect. Dis. Soc.* **7**, 129–135 (2018).
72. Saria, S., Rajani, A. K., Gould, J., Koller, D. & Penn, A. A. Integration of early physiological responses predicts later illness severity in preterm infants. *Sci. Transl. Med.* **2**, 48ra65 (2010).
73. Sullivan, B. A. et al. Early pulse oximetry data improves prediction of death and adverse outcomes in a two-center cohort of very low birth weight infants. *Am. J. Perinatol.* **35**, 1331–1338 (2018).
74. Moorman, J. R. et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. *J. Pediatr.* **159**, 900–906.e1 (2011).
75. Fairchild, K. D. et al. Endotoxin depresses heart rate variability in mice: cytokine and steroid effects. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **297**, R1019–R1027 (2009).
76. Fairchild, K. D. et al. Septicemia mortality reduction in neonates in a heart rate characteristics monitoring trial. *Pediatr. Res.* **74**, 570–575 (2013).
77. James, C. A., Wachter, R. M. & Woolliscroft, J. O. Preparing clinicians for a clinical world influenced by artificial intelligence. *JAMA* **327**, 1333–1334. <https://doi.org/10.1001/jama.2022.3580> (2022).
78. Emanuel, E. J. & Wachter, R. M. Artificial intelligence in health care: will the value match the hype? *JAMA* **321**, 2281–2282 (2019).
79. Swanson, J. R. et al. Neonatal intensive care unit length of stay reduction by heart rate characteristics monitoring. *J. Pediatr.* **198**, 162–167 (2018).
80. Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E. & Valley, T. S. Racial bias in pulse oximetry measurement. *N. Engl. J. Med.* **383**, 2477–2478 (2020).
81. Andrist, E., Nuppau, M., Barbaro, R. P., Valley, T. S. & Sjoding, M. W. Association of race with pulse oximetry accuracy in hospitalized children. *JAMA Netw. Open* **5**, e224584 (2022).
82. Vesoulis, Z., Tims, A., Lodhi, H., Losos, N. & Whitehead, H. Racial discrepancy in pulse oximeter accuracy in preterm infants. *J. Perinatol.* **42**, 79–85 (2022).
83. Sullivan, B. A. & Keim-Malpass, J. BARRIERS to early detection of deterioration in hospitalized infants using predictive analytics. *Hosp. Pediatr.* **11**, e195–e198. <https://doi.org/10.1542/hpeds.2020-004382> (2021).
84. Winters, B. D. et al. Technological distractions (part 2): a summary of approaches to manage clinical alarms with intent to reduce alarm fatigue. *Crit. Care Med.* **46**, 130–137 (2018).
85. Joshi, R. et al. Pattern discovery in critical alarms originating from neonates under intensive care. *Physiol. Meas.* **37**, 564–579 (2016).
86. Escobar, G. J. et al. Automated identification of adults at risk for in-hospital clinical deterioration. *N. Engl. J. Med.* **383**, 1951–1960 (2020).
87. Keim-Malpass, J. et al. Advancing continuous predictive analytics monitoring: moving from implementation to clinical action in a learning health system. *Crit. Care Nurs. Clin. North Am.* **30**, 273–287 (2018).
88. Ghassemi, M., Oakden-Rayner, L. & Beam, A. L. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit. Health* **3**, e745–e750 (2021).
89. Fairchild, K. D. & O'Shea, T. M. Heart rate characteristics: physiologic markers for detection of late-onset neonatal sepsis. *Clin Perinatol* **37**, 581–598 (2010).
90. Gur, I. et al. A mathematical algorithm for detection of late-onset sepsis in very-low birth weight infants: a preliminary diagnostic test evaluation. *Indian Pediatr* **51**, 647–650 (2014).
91. Mithal, L. B., Yogeve, R., Palac, H., Gur, I. & Mestan, K. K. Computerized vital signs analysis and late onset infections in extremely low gestational age infants. *J Perinat Med* **44**, 491–497 (2016).
92. Cabrera-Quiros, L. et al. Prediction of Late-Onset Sepsis in Preterm Infants Using Monitoring Signals and Machine Learning. *Crit. Care Explor* **3**, e0302 (2021).
93. Masino, A. J. et al. Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. *PLoS One* **14**, e0212665 (2019).

## AUTHOR CONTRIBUTIONS

B.A.S. wrote the first draft of the manuscript. All authors contributed significant input with editing and content.

## FUNDING

Eunice Kennedy Shriver National Institute of Child Health and Human Development: K.F.: HD072071 and B.S.A.: HD097254.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Brynne A. Sullivan.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.