Universiteit Utrecht

Developing innovative methods for measuring body temperature in preterm infants to enhance prediction of late-onset sepsis

Abstract

Background and aim: Significant progress has been made in employing machine learning algorithms to predict late-onset sepsis (LOS). Despite the availability of body temperature measurements, it is underutilized due to external influences like incubator temperature. This study aimed to develop new methods to measure body temperature.
Methods: In this retrospective cohort study, preterm infants (GA < 32 weeks) from the Wilhelmina Children's Hospital (WKZ) were included. Patients were divided into LOS or control groups based on blood culture results. Body and incubator temperatures were extracted around the time a positive blood culture collection, and equivalent timestamps for controls. Two methods were evaluated at five time points: at blood culture collection (t=0), two hours after (t=2), and twenty-four (t=-24), four (t=-4), and two (t=-2) hours before. Firstly, the absolute median difference over the past 30 minutes was assessed for each time point. The second method focused on the disparity between body and incubator temperatures. Differences between group were tested using the Wilcoxon singed-rank test.
Results: After matching, two groups of 362 patients were analysed. The MAD showed significant differences at t=0 and t=2. The body-incubator temperature difference showed significant results at t=-2, t=0 and t=2.
Conclusion: Both methods demonstrated differences in body temperature measures between LOS and control groups at various time points, indicating their potential for integrating body temperature into machine learning algorithms.

Keywords: Late-onset sepsis | Body temperature | Incubator temperature | Variability

Max Botman (6798632)
Date: 28/06/2024
Master Thesis Applied Data Science, Utrecht University

1. Introduction

Neonatal sepsis is still a significant contributor to the global neonatal mortality rate (Fleischmann et al., 2021). It has been estimated that there is a global incidence rate for neonatal sepsis of approximately 2.200 per 100.000 births, with mortality rates between 11% and 19% (Fleischmann-Struzek et al., 2018). Sepsis is a systemic, life-threatening condition as a result of a dysregulated response to an infection, which can lead to multiple organ dysfunction (Popescu et al., 2020). In premature infants, a distinction is made between two types of sepsis: early-onset sepsis (EOS) and late-onset sepsis (LOS). EOS typically occurs within the first 72 hours after birth, whereas LOS manifests after 72 hours (Köstlin-Gille et al., 2021). An infection that leads to EOS is most commonly due to transmission during birth (Fedaa Noah et al., 2022). LOS is mainly caused by nosocomial infection but can also be caused by infections through community environments, such as contact from healthcare workers or caregivers (Coggins & Glaser, 2022). Current research will focus on LOS.

Besides the high mortality rates, LOS can have severe long-term consequences, even with adequate treatment (Sewell et al., 2021). Research has shown that neonatal infections are associated with poor growth and increased risk of neurodevelopmental problems (Cai et al., 2019; Strunk et al., 2014). Those neurodevelopmental problems can consequently lead to neurocognitive impairment of motor, cognitive, language, and behavioural skills (Pek et al., 2020). In addition to its physical implications, LOS can lead to prolonged hospitalization, thereby increasing hospitalization costs (Kaye et al., 2014). Given the high mortality rates and the possible implications after adequate treatment, a timely LOS diagnosis is of utmost importance (Dong & Speer, 2015).

However, diagnosing LOS is a lengthy and challenging process due to several factors. The symptoms are often nonspecific and can mimic other conditions such as inflammatory syndromes (Bethou & Bhat, 2022; Nyenga et al., 2021). Early clinical signs include

worsening respiratory distress, feeding intolerance and temperature instability (Walker et al., 2019). The golden standard for a diagnosis, a positive blood culture, can take up to two days to confirm and may suffer from contamination (Guido et al., 2016; Wagstaff et al., 2019). Due to the high fatality risk of LOS, antibiotics may be started without confirmed positive blood cultures, or even after negative results. This can lead to longer hospital stay, increased mortality and antimicrobial resistance (Grant et al., 2018; Korang et al., 2021; Thomas et al., 2024). Given the importance of diagnosing LOS and the limitations of current methods, it could be feasible to explore alternative techniques. Machine learning algorithms could offer a way to assist healthcare providers with diagnosing by making predictions on available data (Garstman et al., 2023; Sahu et al., 2022). Research has shown that hidden patterns in vital sign data can show the status of an upcoming disease hours before becoming clinically apparent (Kumar et al., 2020; Wiens & Shenoy, 2018). By using the available data and identifying potential underlying patterns, it may become possible to predict LOS earlier than the conventional methods (Meeus et al., 2024).

There have already been numerous studies that looked into the possible applications of machine learning algorithms for predicting LOS (O'Sullivan et al., 2023; Sahu et al., 2022). The results of these studies demonstrate that machine learning models can make a significant contribution to the prediction of LOS. Some models also exhibited a strong performance in early prediction (Cabrera-Quiros et al., 2021; Song et al., 2020). By utilizing vital markers such as heart rate, blood pressure and oxygen saturation, it was possible to make a prediction 5 to 48 hours before the onset of clinical symptoms. More recently, a study from van den Berg et al. (2023) aimed to develop a model that is also applicable within the clinical field, with the aim of making it useful for clinicians. This shows that significant progress is being made in the field of prediction modelling for LOS, but some areas remain underexplored. A major

opportunity may lie in expanding the use of low-resolution vital data, which is often readily available (Chen et al., 2010).

Body temperature is one of these vital markers that is often not incorporated into machine learning models (Verstraete et al., 2015). An important reason for this is that body temperature is often perceived as too nonspecific as a clinical sign (Bekhof et al., 2013). This can be attributed to various reasons. Firstly, temperature exhibits significant variability within and between patients, making it difficult to establish a clear trend (Ahmad et al., 2016; Frazer et al., 2019). Another significant issue with body temperature, is that the temperature of neonatal infants can be significantly influenced by external factors (Kumar et al., 2020). Especially the incubator temperature may pose challenges in obtaining accurate measurements of the actual body temperature (Shah & Padbury, 2014). When changes in body temperature are observed, the incubator temperature can be manipulated to prevent extremely high or low body temperatures of the infant, and stabilize the body temperature (Bekhof et al., 2013). Therefore, it becomes difficult to discern whether the baby's temperature is intrinsic or a result of the incubator's temperature regulation (Bekhof et al., 2013).

Despite the disadvantages associated with measuring body temperature, the inclusion could provide valuable information. First of all, temperature is one of the clinical signs that may indicate the onset of neonatal sepsis (Sullivan & Fairchild, 2022). Due to systemic inflammation, sepsis is often associated with alterations in body temperature, such as fever or hypothermia (Rumbus & Garami, 2018). Additionally, temperature is a vital sign that is relatively easy to measure. Often, body temperature is one of the first parameters that is measured in the neonatal intensive care units (NICU) (Smith, 2014). Another benefit of using body temperature is that the measurement process itself is often not highly invasive for a neonatal infant, especially the data that is routinely collected via sensors, such as in the diaper or axillary (Lei et al., 2021). Given these benefits associated with using body temperature, it

becomes clear that developing a reliable measure of body temperature without the effects of the incubator can be highly valuable.

Recent advancement in the measurement of body temperature have shifted towards different and novel methodologies. One approach focuses on the variability in body temperature rather than fixed thresholds, suggesting it could be a better indicator for infections (Coiffard et al., 2023; Drewry et al., 2013). However, there is no clear consensus in the literature on the direction of this relationship. While some studies suggest an increased temperature variability in LOS patients compared to controls (Bekhof et al., 2013; Bhavani et al., 2019), others propose that LOS may reduce temperature variability (Buchan et al., 2012; Papaioannou et al., 2019). Therefore, it might be valuable to re-examine the direction of this effect. Another approach involves the use of central-peripheral temperature differences as a clinical marker for LOS. This method monitors the difference between the central and peripheral body temperature, where differences exceeding two degrees Celsius over a continuous period of four hours might be a significant indicator for LOS (Leante-Castellanos et al., 2017; Patil et al., 2023; Ussat et al., 2015). This study also demonstrates that if this difference maintains for four hours, it remains resistant to changes in incubator temperature (Leante-Castellanos et al., 2017).

These advancements in temperature measurement are promising. However, limited research has been conducted exclusively on the interaction between body temperature and incubator temperature. A novel approach could involve investigating the difference between body temperature and incubator temperature. The incubator temperature is crucial for maintaining the body temperature of the neonate within a stable range, which is essential for their well-being (Ringer, 2013). To ensure this stability, the incubator temperature is adjusted dynamically to prevent the neonates from experiencing extreme temperatures (Bekhof et al., 2013). While this can pose difficulties because the infant's body temperature responds to the

environmental temperature, making reliable measurements challenging, it could also reveal underlying patters (Hannouch et al., 2020). When there is a significant difference between the infant's body temperature and the set temperature of the incubator, the incubator temperature is adjusted (Jost et al., 2017). For instance, if the infant's body temperature rises considerably, the incubator temperature is lowered to help regulate and lower the infant's body temperature. This could create a temporary large difference between the two temperature measurements, which may occur more frequently in LOS patients, as fluctuations in body temperature are a common symptom of LOS (Cabrera-Quiros et al., 2021). Observing these temporary differences between the two measures could reveal underlying patterns that may serve as a clinical marker for LOS.

In conclusion, integrating body temperature into a machine learning algorithm could be valuable, but is currently not widely practiced due to challenges associated with the measurement of body temperature. Attempts for precise measurement have shown to be promising but have some limitations regarding the influence of the incubator temperature. Despite the drawbacks associated with temperature measurement, its inclusion would nevertheless be a valuable addition to a prediction model. Given these gaps in knowledge, interests and emerging developments, current research sets out to find novel measurements for body temperature. The results of this study can hopefully contribute to a better understanding of the use of body temperature for an early prediction of LOS. Thereby reducing the mortality rates and preventing unnecessary antibiotic use.

2. Methods

2.1 Study design and patients

The patient data was collected from a large cohort of preterm babies from the NICU in the Wilhelmina Children's Hospital (WKZ) between April 2008 and May 2019. Participants included in the cohort were infants with a gestational age ≤ 32 weeks, who were admitted to

the hospital within 48 hours after birth and had a complete medical record from admission until at least 30 days after birth or until discharge. Infants presenting severe congenital syndromes or those who died within the initial 96 hours after birth were excluded from the study sample. A total of 2686 infants met the inclusion criteria and entered the study. Given the nature of the data, current research employs a retrospective study design.

Baseline demographics were collected, such as gender, gestational age, date of death and the birth date-time. In addition, data on the results of the blood cultures was collected, including the result and corresponding timestamp. The timestamp of the blood culture corresponds to the moment the blood culture was collected. For temperature, various metrics have been collected. Firstly, the body temperature has been documented using diverse methodologies, including rectal, skin and axillary measurements. Furthermore, the set temperature and the temperature within the incubator have been systematically monitored. For both body temperature and incubator temperature, data is recorded at most at one-minute intervals for each individual patient. The current study utilizes skin temperature measurements and the measured temperature within the incubator. These variables were chosen based on the comprehensiveness of the data and clinical expertise. While body temperature data is consistently measured by the same device, infants may change incubators over time.

2.2 Patient selection

The patients that remained in the data were split into a LOS or a control group. Patients were assigned to the LOS group when a positive blood culture was obtained between 72 hours and 30 days after birth. Positive blood cultures before 72 hours are classified as EOS and thus are excluded from the experiment. Patients with a negative, absent or positive blood culture but outside the specified time frame were assigned to the control group. Since it is

possible for a patient to get multiple blood cultures taken when hospitalized it has been decided that the first positive blood culture with complete data was used for classification. To ensure a valid comparison between the two groups during the analysis, a matching procedure was utilized. This implies that each patient in the LOS group is matched with one patient in the control group based on specific characteristics. In the current study, the patients were matched based on gestational age (±2 days) and gender. These variables were chosen because literature shows that both body and incubator temperature can differ due to both gestational age and gender (Lyu et al., 2015; Ralphe et al., 2021). The matched control patient is randomly picked from all the control patients with the same gestational age and gender. So in this study, for instance, a LOS patient with a positive blood culture at 100 hours after birth would be matched with the data of a control patient also at 100 hours after birth, with the same values for gestational age and gender. This matching procedure guarantees a certain degree of similarity between the two groups.

2.3 Descriptive statistics analysis

For an overall description of the patients, the data was explored for both patient groups. The descriptive statistics was presented for both the raw data prior to matching and the matched data. The analysis of the raw data was done for all initial included patients. The data regarding both demographic and temperature variables are presented. In instances where the data follows a normal distribution, means and standard deviations are provided. For non-normal distributions, the median and interquartile ranges are indicated. The normality of the data is assessed using the Shapiro-Wilk test. The differences in means for continuous variables for the raw data were tested using a student's t-test for normal distributions and a Mann-Whitney U test for non-normality. Categorical variables were assessed with the chi-square test. For the matched data a student's t-test was used for normally distributed

continuous variables and a Mann-Whitney U test for non-normal distributions. For categorical variables a McNemar's test was used. P-values with alpha less than .05 were considered statistically significant.

2.4 Measurements

To assess novel measurements to measure body temperature, two distinct methods are employed. The first method examines the variability of body temperature over time, whereas the other method investigates the difference between body temperature and incubator temperature over time. Both methods are compared between the LOS group and the control group on five different time points. The first time point that is extracted is the timestamp when a blood culture was taken (t = 0). To be able to determine if this measure could be useful at an earlier stage the differences are also examined twenty-four, four and two hours before the blood culture result (t = -24, t = -4, t=-2). To test if the potential effect persists after the positive blood culture, an additional time point is considered two hours after the blood culture (t = 2).

2.4.1 Variability measurement

Increasing evidence suggests that the focus of making use of temperature should not solely be on values above or below certain boundaries. The variability in the patterns of temperature may be even more specific for an infection (Coiffard et al., 2023; Drewry et al., 2013). The method used in current research to encompass this variability is the Median Absolute Deviation (MAD). This method examines the difference between each data point and the median temperature over a specific time period. The advantage of MAD is that it is a robust method that is resistant to the effects of large outliers. For each data point, the MAD will be calculated over the previous 30 minutes to assess the variability of that time period.

2.4.2 Body-incubator temperature difference measurement

The secondary method is primarily aimed at disentangling the body temperature from the temperature of the incubator. The basic concept is that the body temperature is compared to the incubator temperature for each available data point. One would expect that the difference in those temperatures remain relatively stable over time in the absence of an infection. When an infant's temperature begins to change due to an infection, the magnitude of this difference would change. For example, in the case of increasing body temperature, the incubator temperature would drop, resulting in an increase in the difference between both temperatures compared to previous measures.

For this body-incubator difference method, a few preprocessing steps have been performed. Body temperature exhibits a high degree of variability and may also contain outliers due to factors such as sensors detaching from the body. Since this method focuses more on underlying patterns rather than on the actual variability of body temperature, it was decided to apply a smoothing algorithm. By doing this, the analysis is less influenced by major outliers or significant variability in the data. Smoothing is done for both the body temperature as the incubator temperature by using a moving average algorithm. For the smoothing, a window of 30 minutes was set. After smoothing both time-series, the temperature difference has been calculated for each possible time point by subtracting the incubator temperature from the body temperature. To determine if this disparity is an actual deviation, it is compared with the average difference observed over the previous hour. Since this value can become positive and negative, the absolute value was taken to prevent the values from cancelling each other out.

2.5 Data processing

To compute values for both measures at the different time points, the data is processed in the following manner. For each LOS patient, a subset of the data was extracted to reduce the amount of data to be processed. This subset included data collected 48 hours before to two hours after a positive blood culture for a LOS patient and will be used to extract the five time points of interest. For the control group, the data was extracted and aligned with the timeline of the matched patient in the LOS group. When dealing with body temperature data, it is conceivable that a sensor may become detached, leading to anomalies in the dataset. To prevent the substantial impact of these anomalies on the data, large outliers were transformed. Based on clinical expertise, it was decided to replace all body temperature data points below 35 degrees with the preceding data point.

The next step in the data processing was extracting the incubator data for the patients in both groups. Given that the NICU at the WKZ utilizes various incubators and infants may transition between them over time, all the incubator temperature data from different incubators was merged to construct a comprehensive time series. The incubator temperature that is collected from the large time series aligns with the already disclosed body temperature. So, for each time point in the body temperature time series, the corresponding incubator temperature is joined. In cases where there is no corresponding incubator temperature for a specific body temperature measurement, the data point is replaced by the previous complete measurement. The resulting dataset is a combination of all the time series for both the LOS and control patients.

Based on this complete dataset, values are calculated for both methods as described in section 2.4. For example for the MAD, a score is computed for each patient at every data point over the preceding half hour, resulting in an additional column indicating the MAD score for each measurement. The same procedure is applied to the body-incubator temperature

difference. After computing the values for all data points and patients, the dataset is filtered to include only the five specified time points of interest (t=-24, t=-4 etc.). Consequently, each patient ends up with a single measurement for each of these time points for both methods after data processing. Ultimately, each patient has 10 rows of data, one for each of the five time points for both methods.

2.6 Statistical analysis

To determine the statistical difference between the LOS and the control group on both measures, the means of the groups are compared at the five different time points. The difference between the two groups is assessed using a paired t-test for each time point if the data is normally distributed, otherwise a Wilcoxon signed-rank test will be used. The paired variant of a t-test was chosen because a matching procedure was employed, assuming that the LOS and control patients are thus equivalent. Since each method is tested on five different time points, a total of ten tests are conducted. The statistical analysis has been adjusted accordingly to the matching procedure. A significance level of $\alpha = .05$ was used to determine whether the difference in means was statistically significant. The matching, preprocessing and analysing of the data was performed in in Python 3.11.4 using the packages Pandas 1.5.3, NumPy1.24.3, SciPy 1.10.1, statsmodels 0.14.0.

3. Results

3.1 Baseline characteristics

Based on a positive blood culture, 395 of the infants in the study were classified as a LOS patient. The remaining 2291 patients were assigned to the control groups. During the matching procedure, 33 patients were excluded from the LOS group since there was no complete incubator data for the time period around the positive blood culture. The missingness of data for these patients may be due to the fact that the incubator data was

integrated from multiple datasets, but could also be due to the blood culture collection itself. For all the LOS patients with complete data, a match with one of the control patients was found. This results in a sample of 724 patients (362 LOS and 362 control). The comparison of the baseline characteristics is done for both the data prior to and after the matching procedure.

The results of the baseline characteristics before the matching procedure are shown in Table 1. The baseline data before the matching procedure consists of the raw data for each patient. A Mann-Whitney U test was performed to assess the differences between both groups on gestational age, body temperature and incubator temperature. A chi-squared test was used to examine the difference in gender. The median gestational age differed significantly between the two groups (U = 297710.0, $p < .001$). The median gestational age for the LOS patients was 28.2 weeks; for the control patients, this was 30.2 weeks. The percentage of female babies in the LOS and control group was respectively 48.6% and 46.6%, so both groups did not differ statistically on gender X^2 (1, N = 2686) = 0.60, p = .438. The body temperature for LOS patients (Mdn = 36.22) did not differ significantly from the body temperature of the control patients (Mdn = 36.24), U = 433847, p = .204. When looking at the incubator temperature, the results show that the median temperature for the LOS group was 31.22 degrees Celsius and 31.57 degrees Celsius in the control group indicating that the incubator temperature is significantly higher in the control group (U = 364388, $p < .001$).

Table 1

Baseline characteristics raw data

Raw data	LOS	Control	P-value
	Total (n=395)	Total (n=2291)	
Demographics			
Gestational age (weeks)	28.2 (26.2 – 30.2)	30.2 (28.2 – 31.3)	<.001*
Sex (female), n (%)	192 (48.6)	1062 (46.4)	.439
Temperature			
Body temperature	36.93 (35.93 – 36.45)	36.25 (35.91 – 36.49)	.204
Incubator temperature	31.22 (30.44 – 31.86)	31.57 (30.76 – 32.30)	<.001*

* indicates p<.05

The baseline characteristics for the patients remaining in the data after the matching procedure are presented in Table 2. These statistics are calculated over all five time points that are utilized in the final analysis. A Mann-Whitney U test showed that the LOS group and the control group did not differ significantly on gestational age after the matching procedure (U = 65286, p = .933). A McNemar test was conducted to examine differences in gender distribution between the two groups. The test revealed no significant difference between the two groups $X^2(1, n = 724) = 0.003$, p = .958. Since these variables were used to match patients, this provides an indication that the matching procedure has yielded the desired outcome. When analysing the temperature variables, the results show that the mean difference in body temperature between the LOS group (Mdn = 36.92) and the control group (M = 36.90) was not significant (U=67941, p=.390). The results of the incubator temperature show that the medians of the LOS group (Mdn = 31.92) and the control group (Mdn = 31.99) did not differ significantly (U = 66246, p = .790).

Table 2

Baseline characteristics matched data

Matched data	LOS	Control	P-value
	Total (n=362)	Total (n=362)	
Demographics			
Gestational age (weeks)	28.2 (26.3 – 30.2)	28.3 (26.3 – 30.2)	.933
Sex (female), n (%)	182 (50.3)	182 (50.3)	.958
Temperature			
Body temperature	36.92 (36.72 – 37.15)	36.90 (36.70 – 37.14)	.390
Incubator temperature	31.92 (30.79 – 32.99)	31.99 (30.82 – 32.83)	.790

3.2 Statistical analyses

The data for both methods were not normally distributed at any of the time points; therefore, the difference were tested using a Wilcoxon signed-rank test. The outcomes of all the tests can be found in Table 3. When examining the differences between the two groups in terms of the median absolute difference (MAD), the results show a significant effect for two

time points: t = 0 (W=23133, *p*=<.001) and t=2 (W=26909, *p*=.019). Indicating that on the time point of a positive blood culture and two hour after this time point the two groups differ significantly on the MAD-score. No significant results were found for t=-24(W=28417, *p*=.100) , t=-4 (W=29835, *p*=.580) and t=-2 (W=27943, *p*=.086). When examining the effect of the difference between body temperature and incubator temperature, it can be observed that a significant effect has been found for t=-2 (W=26662, *p*=.002), t=0 (W=21439, *p*=<.001) and t=2 (W=24419, *p*<.001). This indicates that the groups significantly differ on this measure on the moment of a positive blood culture and 2 hours before and after. No significant effect was found for time point t=-24 (W=30486, *p*=.235) and t=-4 (W=31105, *p*=.381). Additionally to this, the results also show that the differences in means of both groups for both metrics increase around the time of the blood culture. In both measures, the largest difference was found at the time a positive blood culture was confirmed.

Table 3

Analysis output for each time point

	Time point	Median LOS	Median Control	W-value	P-value
Median absolute difference					
	t=-24	0.030	0.035	28417	.100
	t = -4	0.029	0.030	29835	.580
	t = -2	0.035	0.030	27943	.086
	t = 0	0.045	0.031	23133	<.001*
	t = 2	0.039	0.025	26909	.019*
Body-incubator difference					
	t=-24	0.111	0.113	30486	.235
	t = -4	0.122	0.103	31105	.381
	t = -2	0.151	0.121	26662	.002*
	t = 0	0.196	0.101	21439	<.001*
	t = 2	0.162	0.107	24419	<.001*

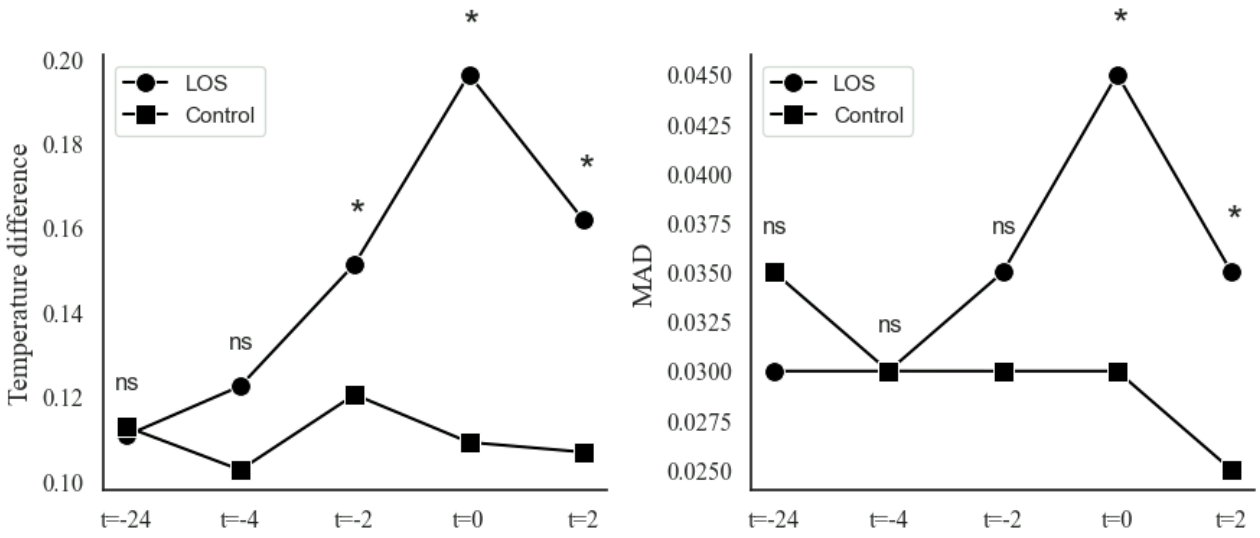
Note: t indicates time before or after blood culture in hours. * indicates p<.05

Figure 1 presents the medians for both groups across both measures at each of the time points. The left figure shows the progression of the medians for both groups over the different time points for the body-incubator temperature difference. In this figure, it is evident that the

differences in the medians become increasingly larger as the time of the blood culture collection approaches (t=0). After the blood culture is taken, the difference remains approximately constant (t=2). The right figure shows the same progression but for the MAD score of the two groups. Notably, the pattern for the MAD-score is different. It begins with a slight difference, followed by a convergence of the medians, and subsequently, the medians differ significantly in the last two time points.

Figure 1

Line plots indicating the medians over time points



Note: ns indicates $p > .05$, * indicates $p < .05$

4. Discussion

The present study aimed to develop a measure that could reduce the influence of external factors, specifically the temperature of the incubator, on body temperature measurements. This is important because body temperature is often neglected in prediction models due to the amount of noise. By measuring body temperature in different ways, it might be possible to achieve even more accurate predictions. To explore such metrics, this study developed two methods. The initial approach focussed on assessing the temporal variability of body temperature, quantified by the MAD. The second method concentrated on the difference

between body temperature and incubator temperature. Both methodologies were examined and tested across various time points.

In the literature, there was a discrepancy regarding the direction of the effect of body temperature variability. While some studies claimed that LOS leads to reduced variability, others demonstrated that LOS results in increased variability in body temperature (Bhavani et al., 2019; Papaioannou et al., 2019). The results found in current research align with the latter. The results indicated that there was a significant difference between the LOS group and the control group for the MAD scores on the last two time point (t=0, t=2). For those time points, the median of the patients in the LOS group was higher than for the patients in the control group. This indicates that patients with LOS demonstrate significantly greater variability in their body temperature compared to the control patients. The time points twenty-four, four and two hours before the blood culture did not show a significant difference between the two groups. Although the significant results can be a promising indication, a critical remark should also be made. It might be challenging to explain the underlying reason for the differences at t=0 and t=2. On one hand, this could stem from actual variability due to the infection. On the other hand, it might also result from the effect of the blood culture collection process itself. Taking the blood culture and opening the incubator could have an effect on the body temperature of the neonates with LOS.

The underlying concept for the second measure, body-incubator temperature difference, was based on the idea that there should be a balance maintained between both temperatures. When this balance is disrupted by a rising or falling body temperature, the incubator temperature is adjusted to restore the difference. The results of the body-incubator temperature difference showed that there was a significant difference between the LOS group and the control group for time points t=-2, t=0 and t=2. For these time points, the average of the body-incubator temperature difference was greater within the LOS patients than in the

control group. These results thus support the notion that in sepsis patients, there is a shift in the balance between body temperature and incubator temperature at the moment a blood culture is taken and at specific time points around this moment. This could be due to changes in both body temperature and incubator temperature. Despite the significant results it remains challenging to demonstrate the clinical relevance. At the time a blood culture is taken, there already is clinical suspicion of LOS. Thus, while the significant results may be innovative, it may yet not be suitable for a timely prediction of LOS. To contribute meaningfully, it is essential that the difference between the two groups are evident twenty-four or more hours before suspicion of LOS.

Based on the current literature research, this is one of the first studies to specifically examine the individual role of incubator temperature on body temperature. The present study may offer novel and valuable insights into the feasibility of obtaining an accurate measurement of body temperature with less interference from the incubator temperature. The discovery of significant results may imply the presence of an underlying pattern that varies between the two groups. These patterns may eventually aid in accurately predicting the likelihood of LOS in neonates, facilitating early interventions and improved patient outcomes. Given the novelty of this method, making comparisons with other study outcomes can be challenging. One study that partially accounted for the effect of incubator temperature is the research conducted by Leante-Castellanos et al. (2017). This study primarily investigated the central-peripheral temperature difference, but also looked into the resistance of this measure to changes in the incubator temperature. Despite the primary focus being on other aspects, the study did reveal comparable patterns between body temperature and incubator temperature.

Even though significant results were found, some limitations should also be addressed. Firstly, the current study only examined the average difference of the two measures between the two groups. The potential contribution of both measures to the accurate prediction of LOS

has not been investigated. The finding of a significant difference at various time points for the two measures does not necessarily imply predictive value. To derive actual clinical value from the measures examined in the current study, it is crucial to evaluate their performance within a machine learning algorithm. If the results indicate that one or both measures are strong predictors of LOS, it would be feasible to extend the existing algorithms. Building on this, it is challenging to assess the clinical relevance of both measures based on the current results. For clinical relevance, it is important to observe differences in patterns before there is a suspicion of LOS. The current results only show differences up to two hours before the collection of a blood culture. A valuable extension of this study would be to investigate whether underlying differences can be detected at an earlier stage.

A third limitation of the study is that each LOS patient is only matched to one control patient. A consequence of this matching procedure, a considerable amount of data from the control group is discarded. Although the matching procedure reduces the influence of confounding variables such as gestational age and gender, the patients who are not included may still contain important information. Additionally, this one-to-one matching procedure could pose a problem regarding the classification of sepsis patients. The current study classified a patient as a LOS patient only when the result of the blood culture was positive. This means that culture-negative patients who could still have had LOS were placed in the control group. This could have resulted in a culture-positive LOS patient being matched with a culture-negative sepsis patient from the control group. Since culture-negative sepsis patients are quite common and would likely exhibit the same underlying pattern in the data, this could introduce some bias in the results (Jiang et al., 2020).

Despite the limitations of the study, it still might provide a foundation for future research. The results obtained in the current study may offer a novel perspective on the utilization of body temperature in LOS prediction. In future research, it is crucial to test the

performance of the measures from this study in predictive algorithms. This will help validate their utility and accuracy in forecasting LOS. Furthermore, it could be valuable in future research to include culture-negative septic patients in the LOS group. This, combined with a matching procedure where each LOS patient can be matched to multiple control patients, could provide an even better understanding of the underlying temperature patterns for both groups.

5. Conclusion

LOS is a significant cause of mortality within the NICU. Due to nonspecific clinical symptoms and the lack of a rapid and reliable test for LOS, early and sometimes unnecessary prescription of antibiotics is still common. There is a need for a new and stable method to predict LOS. In recent years, there have been promising results achieved using machine learning models based on vital data. Despite the advancements in this field, body temperature is not frequently utilized, possibly due to the effects of the incubator temperature. However, given the fact that body temperature is often one of the first data that is measured and is usually non-invasive for a baby, it could be of great value to still be able to incorporate body temperature in such machine learning models.

In this study, two novel methods were developed for a better body temperature measurement in preterm infants. Significant differences were found for both the variability in body temperature and the difference in body-incubator temperature for multiple time points. Both methods might be promising to get a better and more reliable measure for body temperature. However, it is important that further research is conducted based on the limitations discussed. Based on that, it can be assessed whether this approach is indeed a robust measure for distinguishing LOS based on body temperature. When this is the case, it can be incorporated into existing algorithms that might contribute to better predictions for early LOS detection. With those improved predictions, it would be possible to take steps

toward a reduction in the mortality rates in neonates with LOS and prevent unnecessary use of antibiotics.

6. References

Ahmad, M. S., Ali, N., Mehboob, N., Mehmood, R., Ahmad, M., & Wahid, A. (2016). Temperature on admission among cases of neonatal sepsis and its association with mortality. *J Pak Med Assoc*, *66*(10).

Bekhof, J., Reitsma, J. B., Kok, J. H., & Van Straaten, I. H. L. M. (2013). Clinical signs to identify late-onset sepsis in preterm infants. *European Journal of Pediatrics*, *172*(4), 501–508. https://doi.org/10.1007/s00431-012-1910-6

Bethou, A., & Bhat, B. V. (2022). Neonatal Sepsis—Newer Insights. *Indian Journal of Pediatrics*, *89*(3), 267–273. https://doi.org/10.1007/s12098-021-03852-z

Bhavani, S. V., Carey, K. A., Gilbert, E. R., Afshar, M., Verhoef, P. A., & Churpek, M. M. (2019). Identifying Novel Sepsis Subphenotypes Using Temperature Trajectories. *American Journal of Respiratory and Critical Care Medicine*, *200*(3), 327–335. https://doi.org/10.1164/rccm.201806-1197OC

Buchan, C. A., Bravi, A., & Seely, A. J. E. (2012). Variability Analysis and the Diagnosis, Management, and Treatment of Sepsis. *Current Infectious Disease Reports*, *14*(5), 512–521. https://doi.org/10.1007/s11908-012-0282-4

Cabrera-Quiros, L., Kommers, D., Wolvers, M. K., Oosterwijk, L., Arents, N., van der Sluijs-Bens, J., Cottaar, E. J. E., Andriessen, P., & van Pul, C. (2021). Prediction of Late-Onset Sepsis in Preterm Infants Using Monitoring Signals and Machine Learning. *Critical Care Explorations*, *3*(1), e0302. https://doi.org/10.1097/CCE.0000000000000302

Cai, S., Thompson, D. K., Anderson, P. J., & Yang, J. Y.-M. (2019). Short- and Long-Term Neurodevelopmental Outcomes of Very Preterm Infants with Neonatal Sepsis: A Systematic Review and Meta-Analysis. *Children*, *6*(12), Article 12. https://doi.org/10.3390/children6120131

Chen, W., Dols, S., Oetomo, S. B., & Feijs, L. (2010). Monitoring body temperature of newborn infants at neonatal intensive care units using wearable sensors. *Proceedings of the Fifth International Conference on Body Area Networks*, 188–194. https://doi.org/10.1145/2221924.2221960

Coggins, S. A., & Glaser, K. (2022). Updates in Late-Onset Sepsis: Risk Assessment, Therapy, and Outcomes. *NeoReviews*, *23*(11), 738–755. https://doi.org/10.1542/neo.23-10-e738

Coiffard, B., Merdji, H., Boucekine, M., Helms, J., Clere-Jehl, R., Mege, J.-L., & Meziani, F. (2023). Changes in Body Temperature Patterns Are Predictive of Mortality in Septic Shock: An Observational Study. *Biology*, *12*(5), Article 5. https://doi.org/10.3390/biology12050638

Dong, Y., & Speer, C. P. (2015). Late-onset neonatal sepsis: Recent developments. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *100*(3), F257–F263. https://doi.org/10.1136/archdischild-2014-306213

Drewry, A. M., Fuller, B. M., Bailey, T. C., & Hotchkiss, R. S. (2013). Body temperature patterns as a predictor of hospital-acquired sepsis in afebrile adult intensive care unit patients: A case-control study. *Critical Care*, *17*(5), R200. https://doi.org/10.1186/cc12894

Fedaa Noah, N., Leen Jamel, D., & Oday, J. (2022). Perinatal Risk Factors and Early Onset of Neonatal Sepsis. *International Journal of Pediatric Research*, *8*(1). https://doi.org/10.23937/2469-5769/1510088

Fleischmann, C., Reichert, F., Cassini, A., Horner, R., Harder, T., Markwart, R., Tröndle, M., Savova, Y., Kissoon, N., Schlattmann, P., Reinhart, K., Allegranzi, B., & Eckmanns, T. (2021). Global incidence and mortality of neonatal sepsis: A systematic review and

meta-analysis. *Archives of Disease in Childhood*, *106*(8), 745–752. https://doi.org/10.1136/archdischild-2020-320217

Fleischmann-Struzek, C., Goldfarb, D. M., Schlattmann, P., Schlapbach, L. J., Reinhart, K., & Kissoon, N. (2018). The global burden of paediatric and neonatal sepsis: A systematic review. *The Lancet Respiratory Medicine*, *6*(3), 223–230. https://doi.org/10.1016/S2213-2600(18)30063-8

Frazer, J. S., Barnes, G. E., Woodcock, V., Flanagan, E., Littlewood, T., Stevens, R. J., Fleming, S., & Ashdown, H. F. (2019). Variability in body temperature in healthy adults and in patients receiving chemotherapy: Prospective observational cohort study. *Journal of Medical Engineering & Technology*, *43*(5), 323–333. https://doi.org/10.1080/03091902.2019.1667446

Garstman, A. G., Rodriguez Rivero, C., & Onland, W. (2023). Early Detection of Late Onset Sepsis in Extremely Preterm Infants Using Machine Learning: Towards an Early Warning System. *Applied Sciences*, *13*(16), Article 16. https://doi.org/10.3390/app13169049

Grant, C. H., Arnott, A., Brook, T., Horne, A., Hurst, W., Kelly, S., Lang, C., Payne, M., Pert, H., Sparrow, S., Dokubo, P. A., Bee, N., Gibbs, R., & Becher, J.-C. (2018). Reducing Antibiotic Exposure in Suspected Neonatal Sepsis. *Clinical Pediatrics*, *57*(1), 76–81. https://doi.org/10.1177/0009922816689673

Guido, M., Tumolo, M. R., De Donno, A., Verri, T., Serio, F., Bagordo, F., & Zizza, A. (2016). In vitro diagnosis of sepsis: A review. *Pathology and Laboratory Medicine International*, *8*, 1–14. https://doi.org/10.2147/PLMI.S49800

Hannouch, A., Lemenand, T., Khoury, K., & Habchi, C. (2020). Heat and mass transfer of preterm neonates nursed inside incubators—A review. *Thermal Science and Engineering Progress*, *18*, 100553. https://doi.org/10.1016/j.tsep.2020.100553

Jiang, S., Yang, Z., Shan, R., Zhang, Y., Yan, W., Yang, Y., Shah, P. S., Lee, S. K., & Cao, Y. (2020). Neonatal Outcomes Following Culture-negative Late-onset Sepsis Among Preterm Infants. *The Pediatric Infectious Disease Journal*, *39*(3), 232. https://doi.org/10.1097/INF.0000000000002558

Jost, K., Pramana, I., Delgado-Eckert, E., Kumar, N., Datta, A. N., Frey, U., & Schulzke, S. M. (2017). Dynamics and complexity of body temperature in preterm infants nursed in incubators. *PLOS ONE*, *12*(4), e0176670. https://doi.org/10.1371/journal.pone.0176670

Kaye, K. S., Marchaim, D., Chen, T., Baures, T., Anderson, D. J., Choi, Y., Sloane, R., & Schmader, K. E. (2014). Effect of Nosocomial Bloodstream Infections on Mortality, Length of Stay, and Hospital Costs in Older Adults. *Journal of the American Geriatrics Society*, *62*(2), 306–311. https://doi.org/10.1111/jgs.12634

Korang, S. K., Safi, S., Nava, C., Greisen, G., Gupta, M., Lausten-Thomsen, U., & Jakobsen, J. C. (2021). Antibiotic regimens for late-onset neonatal sepsis. *The Cochrane Database of Systematic Reviews*, *2021*(5), CD013836. https://doi.org/10.1002/14651858.CD013836.pub2

Köstlin-Gille, N., Härtel, C., Haug, C., Göpel, W., Zemlin, M., Müller, A., Poets, C. F., Herting, E., & Gille, C. (2021). Epidemiology of Early and Late Onset Neonatal Sepsis in Very Low Birthweight Infants: Data From the German Neonatal Network. *The Pediatric Infectious Disease Journal*, *40*(3), 255. https://doi.org/10.1097/INF.0000000000002976

Kumar, N., Akangire, G., Sullivan, B., Fairchild, K., & Sampath, V. (2020). Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: Big data to the forefront. *Pediatric Research*, *87*(2), 210–220. https://doi.org/10.1038/s41390-019-0527-0

Leante-Castellanos, J. L., Martínez-Gimeno, A., Cidrás-Pidré, M., Martínez-Munar, G., García-González, A., & Fuentes-Gutiérrez, C. (2017). Central-peripheral Temperature Monitoring as a Marker for Diagnosing Late-onset Neonatal Sepsis. *The Pediatric Infectious Disease Journal*, *36*(12), e293–e297. https://doi.org/10.1097/INF.0000000000001688

Lei, D., Tan, K., & Malhotra, A. (2021). Temperature Monitoring Devices in Neonates. *Frontiers in Pediatrics*, *9*. https://doi.org/10.3389/fped.2021.732810

Lyu, Y., Shah, P. S., Ye, X. Y., Warre, R., Piedboeuf, B., Deshpandey, A., Dunn, M., Lee, S. K., & for the Canadian Neonatal Network. (2015). Association Between Admission Temperature and Mortality and Major Morbidity in Preterm Infants Born at Fewer Than 33 Weeks' Gestation. *JAMA Pediatrics*, *169*(4), e150277. https://doi.org/10.1001/jamapediatrics.2015.0277

Meeus, M., Beirnaert, C., Mahieu, L., Laukens, K., Meysman, P., Mulder, A., & Van Laere, D. (2024). Clinical Decision Support for Improved Neonatal Care: The Development of a Machine Learning Model for the Prediction of Late-onset Sepsis and Necrotizing Enterocolitis. *The Journal of Pediatrics*, *266*, 113869. https://doi.org/10.1016/j.jpeds.2023.113869

Nyenga, A. M., Mukuku, O., & Wembonyama, S. O. (2021). Neonatal sepsis: A review of the literature. *Theory and Clinical Practice in Pediatrics*, *3*, 94–101. https://doi.org/10.25082/TCPP.2021.01.006

O'Sullivan, C., Tsai, D. H.-T., Wu, I. C.-Y., Boselli, E., Hughes, C., Padmanabhan, D., & Hsia, Y. (2023). Machine learning applications on neonatal sepsis treatment: A scoping review. *BMC Infectious Diseases*, *23*(1), 441. https://doi.org/10.1186/s12879-023-08409-3

Papaioannou, V. E., Sertaridou, E. N., Chouvarda, I. G., Kolios, G. C., & Pneumatikos, I. N. (2019). Determining rhythmicity and determinism of temperature curves in septic and non-septic critically ill patients through chronobiological and recurrence quantification analysis: A pilot study. *Intensive Care Medicine Experimental*, *7*(1), 53. https://doi.org/10.1186/s40635-019-0267-9

Patil, S. G., Panyang, P. P., Das, N., Biswanath, P., & Baruah, M. N. (2023). Elevated central-peripheral temperature difference in early detection of late-onset sepsis in preterm low birth weight new-born; observational prospective study. *Int J Acad Med Pharm*, *5*(1), 146–151. https://doi.org/10.47009/jamp.2023.5.1.32

Pek, J. H., Yap, B. J., Gan, M. Y., Seethor, S. T. T., Greenberg, R., Hornik, C. P. V., Tan, B., Lee, J. H., & Chong, S.-L. (2020). Neurocognitive impairment after neonatal sepsis: Protocol for a systematic review and meta-analysis. *BMJ Open*, *10*(6), e038816. https://doi.org/10.1136/bmjopen-2020-038816

Popescu, C. R., Cavanagh, M. M. M., Tembo, B., Chiume, M., Lufesi, N., Goldfarb, D. M., Kissoon, N., & Lavoie, P. M. (2020). Neonatal sepsis in low-income countries: Epidemiology, diagnosis and prevention. *Expert Review of Anti-Infective Therapy*. https://www.tandfonline.com/doi/abs/10.1080/14787210.2020.1732818

Ralphe, J. L., Silva, S. G., Dail, R. B., & Brandon, D. H. (2021). The Association Between Very Premature Infant Body Temperatures Over Time and Respiratory Care. *Biological Research for Nursing*, *23*(3), 331–340. https://doi.org/10.1177/1099800420969865

Ringer, S. A. (2013). Core concepts: Thermoregulation in the newborn, part II: Prevention of aberrant body temperature. *NeoReviews*, *14*(5), e221–e226. Scopus. https://doi.org/10.1542/neo.14-5-e221

Rumbus, Z., & Garami, A. (2018). Fever, hypothermia, and mortality in sepsis*. *Temperature: Multidisciplinary Biomedical Journal*, *6*(2), 101–103. https://doi.org/10.1080/23328940.2018.1516100

Sahu, P., Raj Stanly, E. A., Simon Lewis, L. E., Prabhu, K., Rao, M., & Kunhikatta, V. (2022). Prediction modelling in the early detection of neonatal sepsis. *World Journal of Pediatrics*, *18*(3), 160–175. https://doi.org/10.1007/s12519-021-00505-1

Sewell, E., Roberts, J., & Mukhopadhyay, S. (2021). Association of infection in the neonate and long-term neurodevelopmental outcome. *Clinics in Perinatology*, *48*(2), 251–261. https://doi.org/10.1016/j.clp.2021.03.001

Shah, B. A., & Padbury, J. F. (2014). Neonatal sepsis. *Virulence*, *5*(1), 170–178. https://doi.org/10.4161/viru.26906

Smith, J. (2014). Methods and Devices of Temperature Measurement in the Neonate: A Narrative Review and Practice Recommendations. *Newborn and Infant Nursing Reviews*, *14*(2), 64–71. https://doi.org/10.1053/j.nainr.2014.03.001

Song, W., Jung, S. Y., Baek, H., Choi, C. W., Jung, Y. H., & Yoo, S. (2020). A Predictive Model Based on Machine Learning for the Early Detection of Late-Onset Neonatal Sepsis: Development and Observational Study. *JMIR Medical Informatics*, *8*(7), e15965. https://doi.org/10.2196/15965

Strunk, T., Inder, T., Wang, X., Burgner, D., Mallard, C., & Levy, O. (2014). Infection-induced Inflammation and Cerebral Injury in Preterm Infants. *The Lancet Infectious Diseases*, *14*(8), 751–762. https://doi.org/10.1016/S1473-3099(14)70710-8

Sullivan, B. A., & Fairchild, K. D. (2022). Vital signs as physiomarkers of neonatal sepsis. *Pediatric Research*, *91*(2), 273–282. https://doi.org/10.1038/s41390-021-01709-x

Thomas, R., Bijlsma, M. W., Gonçalves, B. P., Nakwa, F. L., Velaphi, S., & Heath, P. T. (2024). Long-term impact of serious neonatal bacterial infections on

neurodevelopment. *Clinical Microbiology and Infection*, *30*(1), 28–37. https://doi.org/10.1016/j.cmi.2023.04.017

Ussat, M., Vogtmann, C., Gebauer, C., Pulzer, F., Thome, U., & Knüpfer, M. (2015). The role of elevated central-peripheral temperature difference in early detection of late-onset sepsis in preterm infants. *Early Human Development*, *91*(12), 677–681. https://doi.org/10.1016/j.earlhumdev.2015.09.007

van den Berg, M., Medina, O., Loohuis, I., van der Flier, M., Dudink, J., Benders, M., Bartels, R., & Vijlbrief, D. (2023). Development and clinical impact assessment of a machine-learning model for early prediction of late-onset sepsis. *Computers in Biology and Medicine*, *163*, 107156. https://doi.org/10.1016/j.compbiomed.2023.107156

Verstraete, E. H., Blot, K., Mahieu, L., Vogelaers, D., & Blot, S. (2015). Prediction Models for Neonatal Health Care–Associated Sepsis: A Meta-analysis. *Pediatrics*, *135*(4), e1002–e1014. https://doi.org/10.1542/peds.2014-3226

Wagstaff, J. S., Durrant, R. J., Newman, M. G., Eason, R., Ward, R. M., Sherwin, C. M. T., & Enioutina, E. Y. (2019). Antibiotic Treatment of Suspected and Confirmed Neonatal Sepsis Within 28 Days of Birth: A Retrospective Analysis. *Frontiers in Pharmacology*, *10*. https://doi.org/10.3389/fphar.2019.01191

Walker, O., Kenny, C. B., & Goel, N. (2019). Neonatal sepsis. *Paediatrics and Child Health*, *29*(6), 263–268. https://doi.org/10.1016/j.paed.2019.03.003

Wiens, J., & Shenoy, E. S. (2018). Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. *Clinical Infectious Diseases*, *66*(1), 149–153. https://doi.org/10.1093/cid/cix731