Late-onset sepsis in preterm infants can be detected preclinically by fecal volatile organic compound analysis: a prospective, multicenter cohort study

Daniel J.C. Berkhout, MD1,2; Britt J. van Keulen, MD1; Hendrik J. Niemarkt, MD, PhD3; Jet R. Bessem, MSc2; Willem P. de Boode, MD, PhD5; Veerle Cossey, MD, PhD5; Neil Hoogenes, BSc2; Christiaan V. Hulzebos, MD, PhD5; Ellen Klaver, MD1; Peter Andriessen, MD, PhD3; Anton H. van Kaam, MD, PhD7; Boris W. Kramer, MD, PhD9; Richard A. van Lingen, MD, PhD10; Aaron Schouten, MSc2; Johannes B. van Gouwodever, MD, PhD11,12; Daniel C. Vlijlbrief, MD, PhD13; Mirjam M. van Weissenbruch, MD, PhD7; Dr. Alfian N. Wicaksono14; Dr. James A. Covington, PhD14; Marc A. Benninga MD, PhD1; Nanne K.H. de Boer MD, PhD15; Tim G.J. de Meij, MD, PhD2

1 Department of Pediatric Gastroenterology, Emma Children’s Hospital / Academic Medical Center, Amsterdam, the Netherlands
2 Department of Pediatric Gastroenterology, VU University Medical Center, Amsterdam, the Netherlands
3 Neonatal Intensive Care Unit, Máxima Medical Center, Veldhoven, the Netherlands
4 Neonatal Intensive Care Unit, Radboud University Medical Center, Radboud Institute for Health Sciences, Amalia Children’s Hospital, Nijmegen, Netherlands
5 Neonatal Intensive Care Unit, University Hospitals Leuven, Leuven, Belgium
6 Neonatal Intensive Care Unit, Beatrix Children’s Hospital/ University Medical Center Groningen, Groningen, the Netherlands
7 Neonatal Intensive Care Unit, VU University Medical Center, Amsterdam, the Netherlands
8 Neonatal Intensive Care Unit, Emma Children’s Hospital / Academic Medical Center, Amsterdam, the Netherlands
9 Department of Pediatrics, Maastricht University Medical Center, Maastricht, the Netherlands
10 Neonatal Intensive Care Unit, Emma Children’s Hospital / Academic Medical Center, Amsterdam, the Netherlands
11 Department of Pediatrics, Emma Children’s Hospital / Academic Medical Center, Amsterdam, the Netherlands
12 Department of Pediatrics, VU University Medical Center, Amsterdam, the Netherlands
13 Neonatal Intensive Care Unit, Wilhelmina Children’s Hospital/ University Medical Center Utrecht, Utrecht, the Netherlands
14 School of Engineering, University of Warwick, Warwick, United Kingdom
15 Department of Gastroenterology and Hepatology VU University Medical Center, Amsterdam, the Netherlands

Key words
Neonatology
Volatile organic compound
Late-onset sepsis
High-Field Asymmetric Waveform Ion Mobility Spectrometry
Electronic nose

Running title
Preclinical detection of sepsis

Correspondence:
Daniel J.C. Berkhout, MD
Department of Pediatrics
VU University medical center
De Boelelaan 1117
1081 HV Amsterdam
The Netherlands
d.berkhout@vumc.nl
tel +31-20-4444444; fax +31-20-4448048

Alternate corresponding author:
Tim G.J. de Meij, MD
Department of Pediatrics
VU University Medical Center
De Boelelaan 1117
1081HV Amsterdam
**Short summary**
Preterm infants developing late-onset sepsis could be discriminated from controls by fecal volatile organic compound analysis. Focusing on individual pathogens, fecal VOCs differed significantly between cases and controls up to three days prior sepsis onset.
Abstract

Background: The intestinal microbiota has increasingly been considered to play a role in the etiology of late-onset sepsis (LOS). We hypothesize that early alterations in fecal volatile organic compounds (VOCs), reflecting intestinal microbiota composition and function, allow for discrimination between infants developing LOS and controls in a preclinical stage.

Methods: In nine neonatal intensive care units in the Netherlands and Belgium, fecal samples of preterm infants born at a gestational age ≤30 weeks were collected daily, up to the postnatal age of 28 days. Fecal VOC were measured by High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS). VOC profiles of LOS infants, up to three days prior to clinical LOS onset, were compared with profiles from matched controls.

Results: In total, 843 preterm born infants (gestational age ≤30 weeks) were included. From 127 LOS cases and 127 matched controls, fecal samples were analyzed by means of FAIMS. Fecal VOCs allowed for preclinical discrimination between LOS and control infants. Focusing on individual pathogens, fecal VOCs differed significantly between LOS cases and controls at all predefined time points. Highest accuracy rates were obtained for sepsis caused by *Escherichia coli*, followed by sepsis caused by *Staphylococcus aureus* and *Staphylococcus epidermidis*.

Conclusions: Fecal VOC analysis allowed for preclinical discrimination between infants developing LOS and matched controls. Early detection of LOS may provide clinicians a window of opportunity for timely initiation of individualized therapeutic strategies aimed at prevention of sepsis, possibly improving LOS-related morbidity and mortality.
Introduction

Despite a decrease in the incidence of late-onset sepsis (LOS) over the past decade, still 34% of all extremely low birth weight and preterm infants, born at a gestational age <29 weeks, experience at least one episode of LOS[1]. Unfortunately, this decline in incidence is not accompanied with a decrease in either LOS associated mortality (12%) or length of hospital stay[1]. Although extensively studied, the exact biological mechanisms of LOS are still poorly understood. Next to indwelling devices, the intestinal microbiota has increasingly been considered to play a pivotal role in LOS etiology. Bacterial dysbiosis, combined with an immature intestinal epithelial barrier and naïve immune system, may increase the risk of transmucosal bacterial translocation, ultimately leading to ‘gut-derived sepsis’. Several microbiota studies have demonstrated microbial alterations preceding the clinical onset of LOS, as compared to controls[2-9]. However, a key feature of these studies is the lack of a consistent LOS-specific microbial signature. Furthermore, implementation of intestinal microbiota analysis in daily clinical practice as an early diagnostic biomarker for LOS is not feasible because of high costs and the complexity to generate and interpret microbiota results in a clinically acceptable fashion.

Fecal volatile organic compound (VOC) analysis may serve as alternative test to monitor changes in microbiota and its metabolic activity. VOCs are carbon-based volatiles, which are considered to reflect microbial composition and function, as they provide insight into the interaction between gut microbiota and host[10, 11]. Fecal VOCs have previously been described to hold potential as noninvasive diagnostic biomarker of diseases that are associated with alterations in intestinal microbiota composition[12, 13]. Recently, in a small proof of principle study, we have shown that fecal VOCs allow for differentiation between preterm infants with and without LOS, up to three days prior to clinical onset of LOS[14]. Since all bacterial species exhibit a unique, metabolic signature[10] and provoke species-specific host-pathogen interactions, fecal VOCs may hypothetically allow for identification of LOS-provoking pathogens at species level.
Primary aim of this study was to investigate whether fecal VOC analysis could discriminate preterm infants developing LOS from controls before the onset of symptoms of LOS, in a large multi-center cohort. The secondary aim was to evaluate whether the discriminative accuracy would increase when focusing on specific causative LOS pathogens.
Methods

Subjects
For a more detailed description we would like to refer you to the supplemental section. In short, between October 2014 and January 2017, infants born at a gestational age ≤30 weeks were consecutively included in this prospective cohort study performed at nine neonatal intensive care units (NICUs) located in the Netherlands and Belgium. The study was approved by all local Institutional Review Boards (protocol A2016.363) and written parental informed consent was obtained from all included patients.

Patient cohort and sample selection
Included preterm infants were identified as LOS cases if all three Vermont Oxford criteria for LOS were met[15]. In case a coagulase-negative Staphylococcus (CoNS) bacteria was isolated from blood culture, infants were allocated to the LOS group only if C-reactive protein (CRP) level was at least once ≥10 mg/L within one week after clinical onset of LOS, in order to minimize the risk of including infants with contaminated blood cultures[16]. Fecal samples obtained at three (t-3), two (t-2) and one (t-1) day(s) prior to clinical LOS onset were used for further VOC analysis.

Each LOS case was strictly matched to one non-LOS control based on the following criteria: center of birth, gestational age, postnatal age, birth weight, days exposed to antibiotics prior to t₀ and enteral feeding type (breastmilk versus formula feeding).

VOC comparisons
Fecal VOC profiles from cases and controls were compared at each pre-defined time point in days (t, t-2 and t-1) and, as extremely preterm infants may not pass stool daily, in an analysis including the last obtained sample prior to t₀. Multiple VOC profile comparisons were performed at these measurement points:

1. In the first analysis, the most frequently cultured species of Gram-positive (except CoNS) (Staphylococcus aureus), Gram-negative (Escherichia coli) and CoNS (Staphylococcus
epidermidis) were identified. Fecal samples from cases and controls were compared at the predefined time points and in an analysis including only the last sample obtained prior to LOS onset. In case both CoNS and a non-CoNS species were obtained from one single blood culture, the LOS case was allocated to the non-CoNS group[15]. Cultures containing two different CoNS species or multiple non-CoNS species in one single blood culture were excluded from further VOC analysis.

2. For the second analysis, pathogens were allocated to either the Gram-positive (except CoNS), Gram-negative or CoNS group, comparing VOC profiles from the last obtained sample prior to t₀ with their corresponding matched control sample. Here, if both Gram-positive and Gram-negative bacteria were obtained from the blood culture, cases were included in both Gram-negative and Gram-positive analysis. Cultures containing CoNS and non-CoNS pathogens were excluded from the CoNS analysis, since CoNS is often considered to be a contaminant in these cases[15].

3. The third analysis included VOC profiles from all LOS pathogens combined and were compared with their matched controls to assess whether the overall VOC signal in LOS differed from controls. Both cross-sectional analysis per time interval (t₃, t₂ and t₁) and an analysis including only the VOC profile of the last fecal sample produced prior to LOS onset was performed.

4. Fourth, to assess the potential influence of center of birth on fecal VOC outcome, we performed three further analyses. Each analysis included only those VOC profiles of the last fecal sample obtained prior to LOS from infants born at one of the three centers who included the highest rates of LOS cases (center 1, 2 and 3).

5. Since blood cultures containing more than one different pathogen are often considered to be contaminated, an additional analysis was performed including only cases with monocultures and compared them with their matched controls.
**VOC analysis by FAIMS technique**

Fecal samples were analyzed using the FAIMS technique (Lonestar®, Owlstone, Cambridge, England) and for a detailed description of the FAIMS technique, we refer to previous studies and the supplemental content[13, 17].

**Statistical analysis**

Demographic and clinical variables were compared using a chi-square, independent t-test or nonparametric test where appropriate. A p-value of <.05 was considered statistically significant. For the VOC-profile analyses, after reduction of the raw data using 2D discrete wavelet transform (a common form of data compression), a Wilcoxon rank-sum test was used to calculate p-values[18]. Subsequently, four classification algorithms, Random Forest, Sparse Logistic Regression, Support Vector Machine, and Gaussian Process Classifier, were used to produce class prediction inside a 10-fold cross validation with 90% of the data as a training set and the remaining 10% as a test set (supplemental content)[19]. ROC curves with corresponding area under the curves (AUCs), sensitivity, specificity, positive predictive value and negative predictive value were produced for each model, selecting the most significant model for notation.
Results

A total of 843 preterm infants were enrolled in this study, of whom 127 both met the criteria for LOS and had sufficient fecal material available for VOC analysis (Figure 1). Patient characteristics and cultured pathogens from the infants excluded based on unavailability of fecal material were similar to the characteristics and pathogens of the current study. In the present study, the sepsis and control groups did not differ with respect to demographic and clinical variables (Table 1). Variation in initial empirical antibiotic treatment between centers is depicted in supplemental table 2. In 14/127 (11%) cases, more than one different pathogen was isolated from the blood culture. Focusing on monocultures, Staphylococcus epidermidis was the most frequently isolated pathogen (n=42), followed by Staphylococcus capitis (n=17), Staphylococcus aureus (n=19) and Escherichia coli (n=11) (Table 2).

Fecal volatile organic compound analysis

An overview of the fecal VOC outcomes is depicted in Table 3, supplemental table 1 and supplemental figure 1. The median date of birth with corresponding interquartile range [IQR] was 5-May-2016 [427 days] for cases and 17-May-2016 [311 days] for controls and could be considered as a reflection of the median sample storage time per study group. Summarized, we observed that fecal VOC profiles of infants with LOS caused by Staphylococcus aureus and Escherichia coli, differed significantly from matched controls infants at all three predefined time points. Corresponding area under the curves (AUCs) in the analysis including only Staphylococcus aureus were .85; .70; .80, and in the analysis including Escherichia coli .88; .99; .86 at t_3, t_2 and t_1 respectively. Pooling the samples obtained most adjacent to t_0 from all Gram-negative (AUC =.77) or Gram-positive pathogens allowed for discrimination between cases and controls. In the subgroup analysis including VOC profiles of all CoNS together, and the analysis focusing on the last obtained sample prior onset from the most frequently cultured species within this CoNS subgroup (Staphylococcus epidermidis), fecal VOCs did not allow for discrimination between LOS cases and controls. In contrast, separating these samples
based on the time points they were obtained, fecal VOCs allowed for discrimination between cases and controls with corresponding AUC of .90; .78; .63 at t-3, t-2 and t-1 respectively.

If all LOS infants were included, irrespective of the cultured pathogen, fecal VOC analysis only allowed for a statistically significant discrimination at t-3. If only the last produced fecal sample prior t0 was included in the analysis, fecal VOC analysis remained statistically significantly different.

Three additional analyses were performed including infants born at one of the three centers, which included the largest number of LOS cases (n=39, n=25, n=16 respectively). Fecal VOCs allowed for statistically significant discrimination in center 1 (AUC±95% confidence interval (CI)), p-value, sensitivity, specificity; .76 [.63-.90], .0005, .76, .72) and center 2 (.86 [.71-1], .0001, .88, .88), but not in center three (.62 [.50-.75], .9693, .74, .59). At least one CoNS was isolated in 33 of 39 (85%) sepsis cases born in center 3, whereas CoNS was identified in 13/25 (52%) cases and 11/16 (69%) cases in respectively center 1 and center 2.
Discussion

In this prospective multicenter cohort study, we investigated the diagnostic potential of fecal VOCs in the preclinical detection of LOS. In general, VOC profiles of 127 infants with LOS could be discriminated from controls before the clinical diagnosis was established. Specifically, LOS caused by *Escherichia coli*, *Staphylococcus aureus* and *Staphylococcus epidermidis* could be differentiated from their matched controls with high predictive value, up to three days before the clinical onset of LOS. Previously, we have demonstrated that fecal VOCs allowed for differentiation between infants developing LOS and matched controls with AUCs of 70.2, 77.7 and 70.4 at respectively three, two and one day prior to LOS onset, but not earlier[14]. However, that study was limited by a wide variety of pathogens and a small sample size (n=36 cases). In addition, VOC analyses in that study were performed by a Cyranose320® eNose (Cyranose320®, Sensigent, US) harboring 32 unique conducting sensors, whereas the currently applied FAIMS technique allows for over 52,000 individual data points[18, 20], based on the mobility of ionized molecules in an electrical field.

In the present study, including a larger number of LOS cases, discriminative accuracies improved compared to the results obtained in the previous study, in case of *Escherichia coli* (n=14) and *Staphylococcus aureus* (n=21) LOS. This suggests the presence of a unique, species-specific fecal VOC profile in LOS. It has been demonstrated, mainly in in-vitro studies, that individual bacterial species are characterized by production of a unique VOC fingerprint[10, 21]. We therefore hypothesized that changes in the composition of the microbiota prior to LOS onset may have largely contributed to the detected alterations in fecal VOC profiles. To date, the intestinal microbiota has increasingly been recognized as an important factor in LOS etiology in preterm infants. Interestingly, an increased abundance of the LOS causing pathogen preceding onset, including *Escherichia coli*[2, 7, 8] and *Staphylococcus aureus*[5, 7], has been demonstrated in several studies.

By pooling both gram-negative pathogens (n=27) and gram-positive pathogens (n=28), discriminative accuracy slightly decreased compared to the analysis involving unique species. This finding may be
explained by the observation that gram-positive and gram-negative pathogens depict specific metabolic processes, allowing for discrimination between both subgroups by VOC analysis[22].

As expected, including fecal samples obtained most adjacent to $t_0$ from all included LOS infants (n=127) in the analysis, discriminative accuracy decreased even further. This may be explained by the wide variety of different LOS pathogens in this study, each pathogen exhibiting a distinct VOC profile. Presumably, the presence of discriminative VOCs in septic infants resulting from production of local and systemic inflammatory biomarkers allowed for the discrimination between LOS and control infants. Another possible explanation is that absence of VOCs originating from commensals characterizing healthy state in controls, including anaerobic bacteria of the genus *Bifidobacterium*[4, 7], allowed for the discrimination in this specific analysis.

In the current study, focusing on the last sample obtained prior to disease onset, fecal VOCs did not allow for discrimination between sepsis caused by either *Staphylococcus epidermidis* (n=42) or CoNS (n=73) in general, and strictly matched controls. Since CoNS species are considered predominant components of both dermal[23] and intestinal microbiome[24], we hypothesize that, in at least a part of the cases, not intestinal mucosal translocation but skin invasive procedures may have been the source of CoNS bacteremia. Consequently, bacterial dysbiosis in the intestines preceding a LOS episode may not have occurred, hampering discrimination between cases and controls based on fecal VOCs in this particular group. This hypothesis is underlined by a recent study, demonstrating the gastrointestinal tract to be dominated by the LOS causing pathogen prior to onset, except for *Staphylococcus epidermidis*[7]. Interestingly, separating these samples based on the time points they were obtained, fecal VOCs allowed for discrimination between cases with *Staphylococcus epidermidis* sepsis and controls at each individual time-point. It has previously been described that the center of birth has a significant influence on fecal VOC profiles, possibly resulting from center-specific protocols on feeding patterns and choice of antibiotics[25]. This finding was confirmed in the present study, since focusing on VOC profiles of infants from the same center of birth resulted in an increased discriminative accuracy between LOS and controls, the only exception being center 3. We
hypothesized that this apparent discrepancy resulted from the presence of CoNS in the vast majority of the sepsis cases (85%) in that particular center, whereas CoNS species were less frequently isolated from blood culture in the remaining two centers. As described previously, detection of CoNS by fecal VOC analysis seems to be less feasible than other pathogens, possibly by a different site of entry.

In the current study, observations are highly dependent on the selected statistical analytical supervised method. Since the performance of various prediction models have been shown to vary considerably, it has been questioned whether a certain classifier model may have universal applicability in different populations[26]. Further studies are needed to assess which model provides optimal accuracy in different research and clinical settings.

Strengths of the current study are the large number of included cases in this multicenter, prospective designed study with daily collection of samples and detailed collection on clinical variables, allowing for applying a matching procedure of cases and controls. This study also has several limitations. Firstly, differences in storage time existed between the first collected samples and the samples collected at the end of the study period. This difference in storage time may have caused degradation of volatiles and consequently has influenced VOC outcome. However, we believe that these potential effects on presented results are limited, since this accounted for both cases and controls[27]. Secondly, microbiota analyses were not performed in the current study. Consequently, study design did not provide prove for our hypothesis that the LOS causing pathogens is abundantly present in the intestines preceding clinical onset. Third, based on our previous study[14], we have only analyzed fecal samples obtained within three days prior to LOS onset. Possibly, VOC differences are present already more than three days prior to LOS onset, providing a larger window of opportunity to prevent LOS. Fourth, although the eNose instrument allows for bedside and relative inexpensive VOC analysis of complex gaseous mixtures, they do not allow for the identification of individual VOCs. Identification of individual, LOS-specific volatiles by means of chemical analytical techniques could allow for development of a tailor-made eNose instrument, with increased accuracy
compared to the non-primed eNose in the present study. Since the currently applied eNose focuses on the entire spectrum of VOCs and is not specifically designed to detect LOS in a preclinical phase, we hypothesize that including more subjects in the analysis would result in an increase in the number of confounding VOCs, deriving from non-relevant and non-avoidable environmental and host-specific sources. Potentially, this may have caused the decrease in discriminative accuracy in the analysis comparing the last obtained fecal samples prior to $t_0$ from infants with a Staphylococcus aureus sepsis, compared to the analysis including the $t_1$ samples. Future studies should focus on identification of discriminative and species-specific VOCs by means of chemical analytical techniques. This may allow for development of LOS-specific eNose sensors to be applied as a diagnostic tool for preclinical detection of LOS in clinical practice, simultaneously providing information about the causative agent. Daily analysis of fecal samples obtained from the diaper would ultimately allow for the early detection and consequently timely intervention of LOS in preterm infants, eventually resulting in a decrease in LOS-related morbidity and mortality. Hypothetically, these sensors could also be incorporated in an incubator, continuously analyzing the VOCs deriving from the infants. In conclusion, we demonstrated that fecal VOC analysis allowed for the preclinical discrimination between infants developing LOS and matched controls, up to three days prior to LOS onset. In particular, the highly pathogenic *Escherichia coli* and *Staphylococcus aureus* were detectable preclinically with high accuracy. Preclinical detection of LOS may provide clinicians a window of opportunity for timely initiation of individualized therapeutic strategies, for example narrow spectrum antibiotics, aimed at prevention of sepsis and might ultimately decrease overall morbidity and mortality in preterm born infants.
Notes

Funding

This work was supported by unrestricted grants from the Maag Lever Darm Stichting, Landelijke Vereniging van Crematoria (Dr. C.J. Vaillant Fonds), Zeldzame Ziekte Fonds and Christine Bader Stichting Irene Kinderziekenhuis. None of the co-authors received an honorarium, grant or other form of payment for the production of this manuscript

Acknowledgements

Conflicts of interest:

Explanation individual contribution of each author:

Authors' contributions: **DB**: study concept and design, acquisition of data, performing measurements, interpretation of data; drafting of the manuscript, wrote first draft of manuscript. **BvK**: study concept and design, acquisition of data, critical revision of the manuscript. **HN**: study concept and design, acquisition of data, drafting of the manuscript. **JB**: data acquisition, data interpretation, critical revision of the manuscript. **WdB**: study design, acquisition of data, interpretation of data, critical revision of the manuscript. **VC**: study design, acquisition of data, interpretation of data, critical revision of the manuscript. **NH**: data acquisition, data interpretation, critical revision of the manuscript. **CH**: study design, acquisition of data, interpretation of data, critical revision of the manuscript. **EK**: data acquisition, data interpretation, critical revision of the manuscript. **PA**: study design, interpretation of data, critical revision of the manuscript. **AvK**: study design, acquisition of data, interpretation of data, critical revision of the manuscript. **BK**: study design, acquisition of data, interpretation of data, critical revision of the manuscript. **RvL**: study design, acquisition of data, interpretation of data, critical revision of the manuscript. **AS**: data acquisition, data interpretation, critical revision of the manuscript. **JvG**: study design, data acquisition, data interpretation, critical revision of the manuscript. **DV**: study design, acquisition of data, interpretation of data, critical revision of the manuscript. **MvW**: study design, acquisition of data, interpretation of data, critical revision of the manuscript. **AW**: analysis and interpretation of data, drafting of the manuscript. **JC**: analysis and interpretation of data, drafting of the manuscript. **MB**: study design, interpretation of data, critical revision of the manuscript. **NdB**: study concept and design, interpretation of data, critical revision of the manuscript. **TdM**: study concept and design, acquisition of data, analysis and interpretation of data; drafting of the manuscript.

In addition, all authors approved the submitted version of the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
References


Figure legends

Figure 1. Flow-chart of participants in the study