Faecal volatile organic compounds in preterm babies at risk of necrotising enterocolitis

Probert, Chris; Greenwood, Rosemary; Mayor, Arno; Hughes, David; Aggio, Raphael ; Jackson, Rachel Elizabeth; Simcox, Liz; Barrow, Heather; García-Finana, Marta; Ewer, Andrew

DOI:
10.1136/archdischild-2019-318221

Document Version
Peer reviewed version

Citation for published version (Harvard):
Probert, C, Greenwood, R, Mayor, A, Hughes, D, Aggio, R, Jackson, RE, Simcox, L, Barrow, H, García-Finana, M & Ewer, A 2019, 'Faecal volatile organic compounds in preterm babies at risk of necrotising enterocolitis: the DOVE study', *Archives of disease in childhood. Fetal and neonatal edition*, pp. F1-F6.
https://doi.org/10.1136/archdischild-2019-318221

Archives of
Disease in Childhood

Faecal volatile organic compounds in preterm babies at risk of necrotising enterocolitis: the DOVE study

Journal:	*Archives of Disease in Childhood*
Manuscript ID	fetalneonatal-2019-318221.R2
Article Type:	Original article
Date Submitted by the Author:	n/a
Complete List of Authors:	Probert, Chris; University of Liverpool, Greenwood, Rosemary; University Hospitals Bristol NHS Foundation Trust, Mayor, Arno; University of Liverpool Hughes, David; University of Liverpool Aggio, Raphael; University of Liverpool Jackson, Rachel; Birmingham Women's NHS Foundation Trust Hospital, Neonatal Unit Simcox, Liz; Birmingham Women's and Children's NHS Foundation Trust Barrow, Heather; Birmingham Women's and Children's NHS Foundation Trust García-Finana, Marta; University of Liverpool Ewer, Andrew; Birmingham Womens Hospital, Neonatal Unit
Keywords:	necrotizing enterocolitis, preterm infants, volatile organic compounds, longitudinal discriminant analysis.

Faecal volatile organic compounds in preterm babies at risk of necrotising enterocolitis: the DOVE study

CS Probert[1], R Greenwood[2], A Mayor[1], DM Hughes[3], R Aggio[1], R Jackson[4], E Simcox[4], H Barrow[4], M García-Fiñana[3], AK Ewer[4,5]

CP and RG joint first author

Addresses:

[1]Department of Cellular and Molecular Physiology, Institute of Translational Medicine, University of Liverpool, Liverpool, UK L693GE

[2]Department of Research and Innovation, University Hospitals Bristol NHS Foundation Trust, Bristol

[3]Department of Biostatistics, Institute of Translation Medicine, University of Liverpool

[4] Neonatal Unit, Birmingham Women's and Children's Hospital NHS Foundation Trust, Birmingham

[5] Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham UK.

Correspondence to : Prof A K Ewer, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham UK B15 2TT; email a.k.ewer@bham.ac.uk

Word Count:

Abstract 247 words

Main text 2808

Abstract

Background: early diagnosis of NEC may improve prognosis but there are no proven biomarkers.

Objective: to investigate changes in faecal volatile organic compounds (VOCs) as potential biomarkers for NEC.

Design: multicentre prospective study

Settings: 8 UK Neonatal Units

Patients: preterm infants <34 weeks gestation

Methods: daily faecal samples were collected prospectively from 1,326 babies of whom 49 subsequently developed definite NEC. Faecal samples from 32 NEC cases were compared to samples from frequency-matched controls without NEC. Headspace, solid phase micro-extraction gas chromatography/mass spectrometry was performed and VOCs identified from reference libraries. VOC samples from cases and controls were compared using both discriminant and factor analysis methods.

Results: VOCs were found to cluster into 9 groups (factors), three were associated with NEC and indicated the possibility of disease up to 3-4 days before the clinical diagnosis was established. For one factor, a one standard deviation increase increased the odds of developing NEC by 1.6 times; a similar decrease of the two other factors was associated with a reduced risk (OR 0.5 or 0.7, respectively). Discriminant analyses identified five individual VOCs, which are associated with NEC in babies at risk, each with an AUROC of 0.75-0.76, up to four days before the clinical diagnosis was made.

Conclusions: Faecal VOCs are altered in preterm infants with NEC. These data are insufficient to enable reliable cotside detection of babies at risk of developing NEC and further work is needed investigate the role of VOCs in clarifying the aetiology of NEC.

Keywords: necrotising enterocolitis; preterm infants; volatile organic compounds; longitudinal discriminant analysis.

Introduction

Necrotising enterocolitis (NEC) is the commonest gastrointestinal emergency in premature infants, occurring in up to 8% of admissions with to 40% mortality.[1-3] This has not significantly improved over recent decades[3, 4] despite overall improvements in neonatal intensive care and survival[3-5] and it is possible that the incidence and mortality from NEC are increasing as more preterm infants survive.[3-5] In addition, NEC is associated with significant morbidity in up to 50% of survivors; either due to the disease itself[6, 7] or the complications of prolonged parenteral nutrition, surgical procedures and short bowel syndrome.[3, 8] Survivors of severe NEC are more likely to have neurodevelopmental impairment. [3, 9, 10] Despite this, the pathogenesis is uncertain and there is no specific treatment. Conservative therapy is limited to discontinuing enteral feeding, antibiotics and supportive care. In advanced disease, surgery to remove affected bowel is a further option.[3]

Early detection of NEC is vital to limit progression to advanced fulminant disease. However, establishing the diagnosis in the prodrome is difficult because clinical features and investigations are often inconclusive. Diagnostic investigations, such as abdominal radiographs, only identify established disease and in fulminant NEC conservative management is often ineffective at preventing disease progression.

Potential biomarkers for NEC including CRP,[11] calprotectin,[12] breath hydrogen concentration,[13] IL6 and IL8[14] have been investigated, but have a limited role in detecting early disease.[12, 15] The likely aetiology of NEC is a result of defective interactions between the intestinal microbiota and the host's response and it has been suggested that high-throughput sequencing technology may enable the detection of an early diagnosis biomarker.[13] Detection of faecal biochemical changes that predict the onset of NEC would enable the instigation of early treatment measures which may reduce disease severity and limit morbidity and mortality.

Volatile organic compounds (VOCs) have a low boiling point (and high vapour pressure) and readily enter the vapour phase. VOCs are emitted from many bodily fluids and contribute to their smell (e.g. faeces).[16] Faecal VOCs are produced by digestion and/or fermentation of intestinal luminal contents and are influenced by the microbiota and diet.[17] Faecal VOCs have a similar, stable pattern in healthy adults,[18] but there are characteristic changes in VOC patterns in some intestinal diseases.[18-21] We have previously examined faecal VOCs in preterm infants[22] and reported a pilot study that showed a characteristic change in VOC patterns in babies developing NEC. [23] De Meij *et al*. used an electronic nose to study faecal gases, and suggested the output could be used to diagnose NEC.[24]

We aimed to investigate further the pattern of faecal VOCs in a larger, prospective, multi-centre study, comparing VOCs in apparently healthy preterm infants developing NEC with those without NEC.

Methods

Participants

Between November 2011 and October 2013 all newborns <34 weeks gestation (and therefore at greater risk of NEC), admitted to eight UK neonatal units [Birmingham Heartlands, Birmingham Women's, Liverpool Women's, Royal Shrewsbury, Royal Wolverhampton, Sheffield Teaching, University Hospital Coventry and Warwickshire, University Hospital Leicester], were considered for recruitment. Babies were excluded if they were unlikely to survive or had significant gastrointestinal anomalies. No centre used probiotics as routine prophylaxis at the time of study.

Parents of potential recruits were approached by a research midwife who obtained written informed consent. Daily faecal samples were collected, if available, from recruitment until NEC was diagnosed or until discharge. Samples were placed in 7ml glass vials (Fisher, Loughborough, UK) and immediately frozen at -20^{o}C. We have previously shown that there is no significant change in VOCs in samples stored for between 1 week, and 12 months at -20C.[25] Batched samples were transported on dry ice at regular intervals to the research laboratory.

Demographic and clinical data on mode of delivery, feeding, antibiotic exposure, respiratory support and episodes of illness were collected. Babies with possible NEC were identified by the clinical team caring for the baby and the NEC grade was subsequently classified by consensus between that team and the independent study team using modified Bell's criteria.[26] Controls did not develop NEC of any grade. Approximately two control babies were frequency matched to each case of NEC with Bell's criteria IIa or greater (i.e. definite NEC): matching was based on demographic factors including mode of delivery, gestational age at birth, birth weight, sex, age at sampling, mode of feeding and antibiotic exposure. Where possible we identified samples from six days prior to the definitive diagnosis of NEC. For controls, a day equivalent to the NEC diagnosis date was determined, and samples were analyzed from the previous 6 days.

The study was sponsored by University of Birmingham and approved by West Midlands Research Ethics Committee (11/WM/0078).

Laboratory methods

We have previously reported the laboratory method used.[27] 50-100mg of faeces was transferred to 10 ml glass vials (Sigma-Aldrich, Dorset, UK) and analyzed by headspace solid phase micro-extraction (SPME). A PerkinElmer Clarus 500 gas chromatography/mass spectrometry (GC/MS) quadrupole benchtop system (Beaconsfield, UK) was used with a Combi PAL auto-sampler (CTC Analytics, Switzerland); a Zebron ZB-624 GC column with inner diameter 0.25 mm, length 60 m, film thickness 1.4 µm (Phenomenex, Macclesfield, UK); carrier gas was 99.996% pure helium (BOC, Sheffield, UK). Samples were pre-incubated for 30 minutes at 60°C prior to exposure to a pre-conditioned 85µm Carboxen®/Polydimethylsiloxane SPME fiber (Sigma-Aldrich, Dorset, UK). GC oven initial temperature was 40°C, held for 1 minute before ramping to 220°C at a rate of 5°C/min and held for 13 min (total run time - 50 min). Solvent delay was set for the first 6 min. The MS was operated in electron impact ionization EI+ mode, scanning from ion mass fragments 10- 300 m/z with an inter-scan delay of 0.1 sec and a resolution of 1000 at full width at half maximum (FWHM). Helium gas flow rate was 1 ml/min. Instrument sensitivity was determined using 2-pentanone with limit of detection 3 times the signal/noise ratio. This laboratory analysis was conducted between November 2013 and August 2014.

Sample VOCs were identified using Automated Mass Spectral Deconvolution System (AMDIS-version 2.71, 2012) software and the NIST mass spectral library (version 2.0, 2011) in conjunction with the R package Metab.[28] IUPAC names were assigned to VOCs.

Statistical methods

Peak area values for VOCs, obtained from AMDIS, were log-transformed and normalized. There is no widely accepted approach for treatment of values below the detection threshold. In this study, VOC values below the detection threshold were given an abundance of -3 (equivalent to three standard deviations below the mean), which corresponds to <1% for normally distributed data. Within sample analysis of VOCs was performed independent of case/control status. Within the VOCs identified we established patterns using principal component analysis (PCA) to produce a smaller number of variables (factors) which captured most of the data variability. PCA uses the correlation matrix to identify compounds that are strongly correlated to one another and divide them into factors (Table s1-s3 for examples). Each factor can be expressed as a linear combination of the VOCs. Using factors rather than individual VOCs avoided the overfitting problems associated with fitting a model with more predictors than events. A similar approach was followed by de Meij *et al*., [24] with the distinction that here the factors were allowed to be correlated with each other

following a non-orthogonal rotation process (promax rotation power of 4, SPSS version 23). We chose this method to reflect different unknown biological mechanisms by which these factors might contribute to NEC and could be associated with each other. Factors were investigated for biological plausibility by regressing against demographic variables. All regression analysis containing multiple data points per infant used robust standard errors in Stata (version 14) and account for the longitudinal nature of the data by clustering by infant. Logistic regression assessed the ability of each factor to predict NEC.

A recently-developed longitudinal discriminant analysis (LoDA) method[29-31] assessed the association of individual VOCs with the risk of developing NEC. Linear mixed models were fit separately to each VOC for both NEC and control samples adjusting for age, gestation, delivery type, sex and a random effect intercept. Discriminant analysis was used for each VOC to determine whether the patient should be classified as at risk of NEC: 1000 bootstrap datasets were generated to evaluate the sensitivity, specificity and area under the curve (AUC). Mean lead time was calculated as the mean time before diagnosis when the infant was correctly scored at risk of NEC.

The classification scheme (Figure 1) was designed to predict the risk of NEC. The threshold chosen was determined by receiver operating characteristic (ROC) curve analysis. When the data showed a baby's risk of developing NEC was above the chosen threshold they were classified as an NEC case (if this were conducted in real time this would allow a clinician to consider early intervention). Whilst a baby's risk remained below the threshold, they were provisionally classified as not developing NEC and their risk were reassessed with the addition of the next day's data.

Results

1,326 premature babies were recruited. 49 babies were diagnosed with NEC Bell's stage IIa or greater and 32 of these had adequate sample collection for analysis; these were matched with 70 controls. There was no evidence of concomitant sepsis in any of the cases and all the controls were deemed healthy at the time of sampling. Neonates did not defaecate daily: a median of 2 samples were analyzed for both cases and controls. We collected 81 samples from cases and selected 163 from controls within the 6-day time window (Supplementary Table S4) following the matching procedure described above. See Table 1 for demographic data.

Table 1 Summary of demographic and clinical factors

	NEC	Control	Test	Probability
	32	70		
Sex			Chi-sq	0.05
Male (n[%[])	13 [41]	43 [61]		
Birth weight (median [IQR], g	795 [976]	1045 [541]	Mann Whitney	0.0002
Mode of delivery			Chi-sq	0.160
Vaginal (n [%})	15 [47]	27 [39]		
Missing data	0	1 [1.4]		
Feeding pattern (n[%])			Chi-sq	0.65
Breast+formula	25 [78]	52 [74]		
Exclusive breast	6 [19]	13 [19]		
Exclusive formula	1 [3]	5 [7]		
Postnatal age at first sample (median [IQR]), d	21 [22]	22 [27]	Mann Whitney	0.66
Duration of gestation (median,[IQR]), w	26.7[2.9]	28.4 [3.8]	Mann Whitney	0.002
Post conception at first sample (median,[IQR]), w	30.6[4.4]	32.4[3.7]	Mann Whitney	0.02
AB use before sampling	32	70		
Bell's grade				
2	14	NA		
3	18	NA		

Number of samples		
Days before diagnosis	**Control**	**NEC**
1	34	16
2	22	8
3	42	20
4	18	14
5	31	16
6	19	8

Number of samples		
Distribution	**Control**	**NEC**
Median	2	2
Q1	1.5	2
Q3	3.5	3
IQR	2	1

50 headspace VOCs were identified (prevalence 45-99%). On average NEC cases had 35.8 VOCs identified per sample and controls 36.4. Two VOCs with similar structures and retention times (30.73 and 30.90 minutes, respectively) were difficult to resolve and their data were combined.

We identified 9 different factors (groups of VOCs) that had correlated abundances and thus might represent a common metabolic (examples in Supplementary Tables S1-3.) or demographic factor (Table 2). They were investigated for biological plausibility and 7 of the 9 were associated with demographic variables linked to NEC: birthweight (factors 1, 6 and 9), formula feed (factor 2 and 9), postnatal age (factor 4), gender (factor 3) and postconception age (factors 5 and 8), biochemical factors (factor 7). Multiple logistic regression identified three predictors of NEC (factors 4, 5 and 8), from which a ROC was generated. A one standard deviation increase in factor 4 increased the chances of NEC (odds ratio (OR) 1.66 [95% confidence interval (CI) 1.13, 2.43]. Increases in factor 5 and 8 were associated with a reduced risk of NEC (OR 0.47 [95% CI 0.30, 0.78] and 0.73 [95% CI 0.53, 0.99]), respectively. Area under the ROC curve was 0.69 [95% CI 0.62, 0.76]. Results separated by day are shown in Table 3.

Table 2 Statistical summary variables (group) observed in the VOC profiles

Group	VOCs in the group (+)=positive direction (-)negative direction	Demographics associated with the group
1	(+) (*E*)-hex-2-enal (+) heptan-2-one (+) heptanal (+) 2-pentylfuran (+) (*Z*)-hept-2-enal (+) dec-1-en-3-one (+) octanal (+) 5-ethylcyclopentene-1-carbaldehyde (+) (*E*)-oct-2-enal (+) nonan-1-ol (+) prop-2-enal (+) propanal (+) 2-ethylfuran (+) pent-1-en-3-one (+) pentanal (+) (*E*)-pent-2-enal (+) hexanal	Birthweight
2	(+) ethanol (+) propan-1-ol (+) butane-2,3-dione (-) ethyl acetate (+) acetic acid (+) propyl acetate (+) propanoic acid (+) propyl propanoate	Post-conceptual age, breast feeding, gender
3	(+) 2,2,4,6,6-pentamethylheptane (+) 2,2,4,4-tetramethyloctane (+) 2,6,10-trimethyl dodecane (+) 3,3,4-trimethyl decane	Not associated with any recorded demographic variables
4	(+) 3-methylsulfanylpropanal (+) benzaldehyde (+) 2-phenylacetaldehyde (+) 2-methylpropanal (-) 3-methylbutanal (-) 2-methylbutanal	Formula milk, age
5	(-) heptan-4-one (+) methanedithione (+) butan-2-one	Post-conceptual age
6	(+) propan-2-one (-) propan-2-ol	Birthweight
7	(+) ethylbenzene (+) xylene (+)(4R)-1-methyl-4-prop-1-en-2-ylcyclohexene	Not associated with any recorded demographic variables
8	(+) 3-methylbutanoic acid (+) 2-methylbutanoic acid (+) butanoic acid	Post-conceptual age
9	(+) 6-methyl-5-hepten-2-one (+) 3,4-dimethylcyclohexan-1-ol (+) 4-acetyl-2,3,4,5,5-pentamethyl2-cyclopenten-1-one	Formula use, birthweight

On average, LoDA was able to predict NEC for each VOC individually up to 4 days before the diagnosis of NEC (e.g., 95% CI for (*Z*)-hept-2-enal = (2.79, 4.42) days). 28 VOCs could classify infants reasonably with an AUC greater than 0.7 (Table 4). The best performing VOCs could identify approximately 70% of NEC cases, and also correctly identified approximately 70% of controls using the cut off which was chosen as the point on the ROC curve closest to the top left corner (the five VOCs with the best level of discrimination are shown in Figure 2).

Table 3 Mean values of the area under the ROC curve in the prodromal period

	Day -6	Day -5	Day -4	Day -3	Day -2	Day -1
AUC	0.70	0.76	0.45	0.70	0.74	0.74
	[0.47, 0.94]	[0.61, 0.91]	[0.24, 0.66]	[0.54, 0.85]	[0.56, 0.93]	[0.59, 0.90]
No. observations	27	47	31	59	30	50

Table 4 Prediction accuracy for VOCs with AUC >.7 using longitudinal discriminant analysis

	Cutoff	Sensitivity	Specificity	AUC	Mean Lead Time (d)
(*Z*)-hept-2-enal	0.37	0.78	0.68	0.76	3.68
pent-1-en-3-one	0.36	0.72	0.73	0.76	3.64
2-ethylfuran	0.38	0.71	0.73	0.76	3.59
pentanal	0.36	0.72	0.73	0.75	3.51
2-pentylfuran	0.37	0.73	0.71	0.75	3.47
dec-1-en-3-one	0.37	0.70	0.73	0.75	3.99
2,6,10-trimethyl dodecane	0.39	0.71	0.72	0.74	4.04
(*E*)-oct-2-enal	0.36	0.72	0.71	0.74	3.76
2-methylbutanoic acid	0.40	0.74	0.67	0.74	3.64
2,2,4,6,6-pentamethylheptane	0.40	0.69	0.74	0.74	3.91
nonan-1-ol	0.37	0.72	0.69	0.74	3.88
2-methylpropanal	0.35	0.71	0.70	0.73	3.70
3-methylbutanoic acid	0.39	0.71	0.67	0.73	3.61
3-methylsulfanylpropanal	0.38	0.68	0.71	0.73	3.90
butan-2-one	0.38	0.72	0.68	0.72	3.62
butanoic acid	0.35	0.71	0.70	0.72	3.94
propyl acetate	0.37	0.74	0.65	0.72	4.00
heptanal	0.36	0.70	0.69	0.72	3.58
propan-1-ol	0.36	0.69	0.70	0.72	3.96
2,2,4,4-tetramethyloctane	0.40	0.68	0.72	0.72	3.93
ethanol	0.37	0.67	0.73	0.72	4.09
(4R)-1-methyl-4-prop-1-en-ylcyclohexene	0.39	0.66	0.71	0.72	3.73
(*E*)-hex-2-enal	0.36	0.68	0.70	0.71	3.64
(*E*)-pent-2-enal	0.37	0.71	0.67	0.71	3.65
propan-2-one	0.35	0.71	0.66	0.71	3.48
octanal	0.35	0.70	0.67	0.71	3.55
propan-2-ol	0.37	0.68	0.66	0.71	3.54
2-phenylacetaldehyde	0.37	0.68	0.69	0.71	4.07

Discussion

We have undertaken GCMS analysis of faecal headspace gases obtained from premature infants at risk of NEC in the largest prospective study of its kind. Over 7000 samples were collected from 1326 babies and pre-diagnosis samples from the 32 babies who subsequently developed NEC were identified and produced 9 distinct groups of VOCs (factors). Comprehensive analysis of VOCs found three factors (4, 5 and 8) that were associated with NEC.

We have previously shown, in a much smaller study of only 6 NEC cases using a less sophisticated analysis, that changes in faecal VOC patterns may occur in the pre-clinical phase of NEC.[23] This finding has also been demonstrated by using an eNOSE system,[24] however the present study included almost three times as many cases of NEC than the study by de Meij *et al*.[24] and provided detailed analysis of individual VOCs (rather than just changes in the pattern) which allowed us to explore, for the first time, the biological plausibility of the changes in VOCs identified.

This biological plausibility, shown by the association of the different factors with demographic variables, may represent changes of VOCs produced by biological processes, such as the colonisation of the gut by specific bacteria. The association of three of the factors with post-conceptual age rather than gestational age or postnatal age alone, suggests an interaction between the maturation of the gut and bacteria colonisation, which may be protective.

Matching was carried out to take account of all demographic variables considered likely to confound the results of the analysis: thus, priority was given to gestation, mode of delivery, place of birth and feed type. Despite this adjustment there remained differences in birthweight, sex and gestation (Table 1) and perfect matching was not achievable as they were matched to available samples at the time that they were suspected of having NEC, to prevent excess frozen samples deteriorating prior to analysis. Nevertheless, the linear mixed models involved in the longitudinal discriminant analysis adjusted for a number of demographic and baseline factors, including age at sampling, gestational age at birth, mode of delivery and sex.

In the sample analysis of different factors identified within the compounds, we observed that of the 9 factors identified following PCA, at least three factors were associated with birthweight, but none of these was also associated with the diagnosis of NEC. Factor 4 contained six correlated VOCs which are associated with formula milk feeds and age (Table 2 and Table s1). The risk of NEC increases with the value of Factor 4, thus the VOCs may be biomarkers of NEC. Three of VOCs in Factor 4 are methyl-aldehydes: 3-methylbutanal and 2-methylbutanal, are specifically related to the metabolism of amino acids, leucine and isoleucine, respectively. Such aldehydes can be converted

into their respective acids, 3-methyl and 2-methylbutanoic acid by aldehyde dehydrogenase. These two acids are present in Factor 8 which is linked to a decreased risk of NEC. A change in the abundance of bacteria or fungi that produce aldehyde dehydrogenase is a possible explanation for these findings.

Bifidobacteria reduce the likelihood of developing NEC.[32] *In vitro* studies have shown that Bifidobacteria produce 3-methyl and 2-methyl butanoic acid.[32] Thus the occurrence of these VOCs, present in Factor 8, might be a reflection of the abundance of Bifidobacteria in the controls. Lactobacilli and Bifidobacteria are a source of butanoic acid present in Factor 8:[32] thus, the greater abundance of butanoic acid in control infants may be a result of the protective effect of these bacteria.

Both Bifidobacteria and Lactobacilli produce butan-2-one which was found in Factor 5 also negatively associated with NEC. A further VOC in factor 5, methanedithione (carbon disulphide)[33] may also be produced by Bifidobacteria.[34] Carbon disulphide has numerous toxic properties, but may be beneficial as it inhibits NfkB signalling by binding to IKKβ, NEMO and NFkB[33, 34] and thus has anti-inflammatory properties. Bifidobacteria 'secretions' have been shown to reduce NfkB in transwell experiments.[34]

Univariate analysis showed several VOCs with abundances which differed between cases and controls to such an extent that they appeared to act as biomarkers for NEC 3-4 days before diagnosis. Some appeared in the factor analysis, but others were significant although not present in Factors 4, 5 or 8. Notable were propanal, pentanal, hexanal: aldehydes identified in Factor 1. Aldehydes are associated with Crohn's disease[20] and are thought to arise from lipid peroxidation as a consequence of epithelial inflammation. The presence of these compounds may be an earlier indicator of enterocolitis.

Both the LoDA and the factor analysis highlight the fact that a number of VOCs are associated with a high risk of developing NEC. The LoDA approach could be extended to consider multiple VOCs simultaneously, although a greater sample size would be required. Validation in further datasets would be required before the usefulness of individual VOCs as potential biomarkers for classification of NEC could be determined.

Although our study is the largest, multicentre, prospective study of its kind, the overall number of NEC cases and the number of prodromal samples remained relatively small and interpretation of our results needs to reflect this. However, we have demonstrated changes in VOCs between babies developing NEC and those who did not, which may offer further insight into the pathophysiology of

NEC, particularly if some of these are associated with microbial metabolic pathways. Understanding these changes may later form the basis for a diagnostic tool.

Further studies already underway intend to try to replicate our data (MAGPIE study: ISRCTN12554594) and further investigate the role of VOCs in the aetiology of NEC.

What is already known on this topic?

- NEC is a serious condition with a high mortality and morbidity. There are no known biomarkers.
- Early diagnosis may reduce illness severity
- Volatile organic compounds (VOCs) have been suggested as potential biomarkers of NEC

What does this study add?

- This is the largest prospective study investigating potential biomarkers for NEC
- VOCs have moderate sensitivity and specificity for identifying cases of NEC 3-4 days before diagnosis
- The biological plausibility of the VOCs identified sheds further light on potential mechanisms for the development of NEC

Contributions: CSP and AE: study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; obtained funding; study supervision. RG, DMH, MGF: statistical analysis; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; AM, RA: acquisition of data; analysis and interpretation of data; drafting of the manuscript; RJ, ES, HB: acquisition of data; drafting of the manuscript

Conflicts of interest: none

Acknowledgements: this study was funded by The Henry Smith Charity and Action Medical Research. MGF acknowledges the support of the UK EPSRC grant EP/N014499/1. DMH acknowledges the support of a UKRI Innovation Fellowship, funded by the MRC (Research Project MR/R024847/1). We are grateful to the local research teams in the participating centers Imogen Storey, Emma Gould, Clair Finnegan, Prakash Satodia, Jacqui Dalglish, Geraldine Ward, Babu Kumaratne, Sharon Kempson, Denise Kirby, Maggie Doodson, Sanjeev Deshpande, Sarah Kirk, Robert Coombs, Julie Cook, Elaine Boyle, Marie Hubbard, Mark Turner, Patrick McGowan and Bronagh Howell. We thank the many nurses who collected samples of faeces for this study and their parents of recruited babies.

References

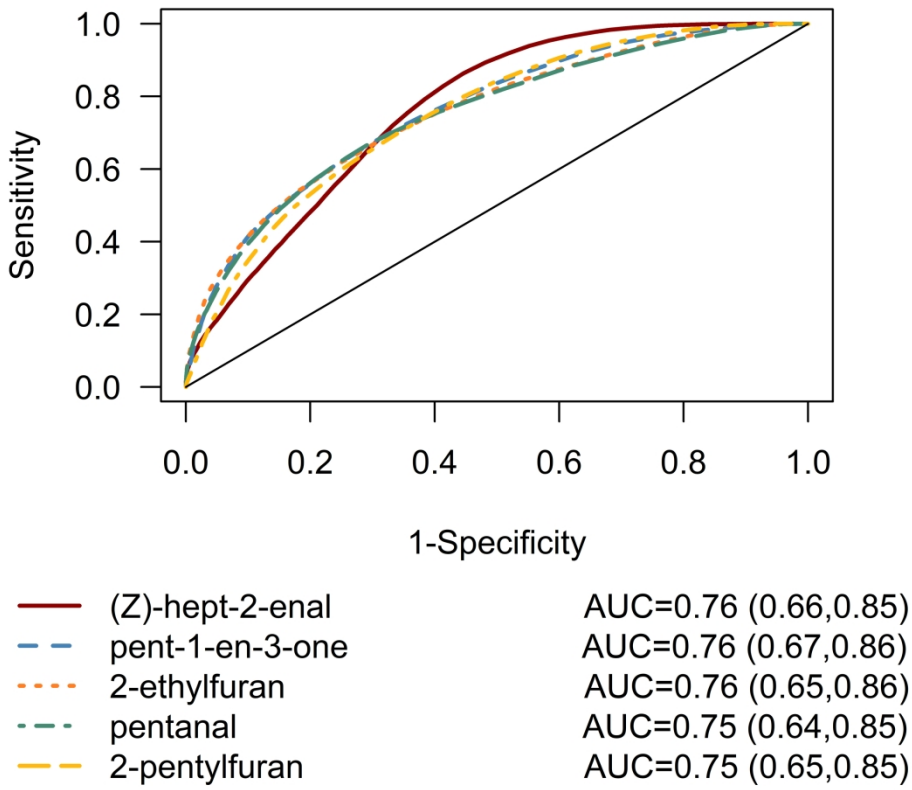
1. Kliegman RM, Walker WA, Yolken RH. Necrotizing enterocolitis - research agenda for a disease of unknown etiology and pathogenesis. Pediatric Research 1993;34:701-708.
2. AM. K. Epidemiology of necrotizing enterocolitis. Acta Paediatr 1994 396:2-7.
3. Yajamanyam PK Rasiah SV, Ewer AK. Necrotizing Enterocolitis: Current Perspectives. Research and Reports in Neonatology 2014;4:31–42.
4. Rees CM ES, Pierro A. Trends in infant mortality from necrotising enterocolitis in England and Wales and the USA Archives of Disease in Childhood - Fetal and Neonatal Edition 2008;93:395-396.
5. Ahle M, Drott P, Andersson RE. Epidemiology and Trends of Necrotizing Enterocolitis in Sweden: 1987-2009. Pediatrics 2013;132:E443-E451.
6. Vohr BR, Wright LL, Dusick AM, et al. Neurodevelopmental and functional outcomes of extremely low birth weight infants in the National Institute of Child Health and Human Development Neonatal Research Network, 1993-1994. Pediatrics 2000;105:1216-1226.
7. Schulzke SM, Deshpande GC, Patole SK. Neurodevelopmental outcomes of very low-birth-weight infants with necrotizing enterocolitis - A systematic review of observational studies. Archives of Pediatrics & Adolescent Medicine 2007;161:583-590.
8. Simon NP. Folow-up for infants with necrotizing enterocolitis Clinics in Perinatology 1994;21:411-424.
9. Hintz SR, Kendrick DE, Stoll BJ, et al. Neurodevelopmental and growth outcomes of extremely low birth weight infants after necrotizing enterocolitis. Pediatrics 2005;115:696-703.
10. Rees CM, Pierro A, Eaton S. Neurodevelopmental outcomes of neonates with medically and surgically treated necrotizing enterocolitis. Archives of Disease in Childhood-Fetal and Neonatal Edition 2007;92:F193-F198.
11. Pourcyrous M, Korones SB, Yang WJ, et al. C-reactive protein in the diagnosis, management, and prognosis of neonatal necrotizing enterocolitis. Pediatrics 2005;116:1064-1069.
12. Carroll D, Corfield A, Spicer R, et al. Faecal calprotectin concentrations and diagnosis of necrotising enterocolitis. Lancet 2003;361:310-311.
13. Young C, Sharma R, Handfield M, et al. Biomarkers for Infants at Risk for Necrotizing Enterocolitis: Clues to Prevention? Pediatric Research 2009;65:91R-97R.
14. Lodha A, Howlett A, Ahmed T, et al. The Role of Interleukin-6 and Interleukin-8 Circulating Cytokines in Differentiating between Feeding Intolerance and Necrotizing Enterocolitis in Preterm Infants. American Journal of Perinatology 2017;34:1286-1292.
15. Josefsson S, Bunn TK, Domellof M. Fecal calprotectin in very low birth weight infants. Journal of Pediatric Gastroenterology and Nutrition 2007;44:407-413.
16. Probert CSJ, Ahmed I, Khalid T, et al. Volatile Organic Compounds as Diagnostic Biomarkers in Gastrointestinal and Liver Diseases. Journal of Gastrointestinal and Liver Diseases 2009;18:337-343.
17. De Preter V, Ghebretinsae AH, Abrahantes JC, et al. Impact of the synbiotic combination of Lactobacillus casei shirota and oligofructose-enriched inulin on the fecal volatile metabolite profile in healthy subjects. Molecular Nutrition & Food Research 2011;55:714-722.
18. Garner CE, Smith S, Costello BD, et al. Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. Faseb Journal 2007;21:1675-1688.
19. Probert CSJ, Jones PRH, Ratcliffe NM. A novel method for rapidly diagnosing the causes of diarrhoea. Gut 2004;53:58-61.
20. Ahmed I, Greenwood R, Costello BD, et al. An Investigation of Fecal Volatile Organic Metabolites in Irritable Bowel Syndrome. Plos One 2013;8.
21. Ahmed I, Greenwood R, Costello B, et al. Investigation of faecal volatile organic

metabolites as novel diagnostic biomarkers in inflammatory bowel disease. Alimentary Pharmacology & Therapeutics 2016;43:596-611.
22. Costello BDL, Ewen R, Ewer AK, et al. An analysis of volatiles in the headspace of the faeces of neonates. Journal of Breath Research 2008;2.
23. Garner CE, Ewer AK, Elasouad K, et al. Analysis of Faecal Volatile Organic Compounds in Preterm Infants Who Develop Necrotising Enterocolitis: A Pilot Study. Journal of Pediatric Gastroenterology and Nutrition 2009;49:559-565.
24. de Meij TGJ, van der Schee MPC, Berkhout DJC, et al. Early Detection of Necrotizing Enterocolitis by Fecal Volatile Organic Compounds Analysis. Journal of Pediatrics 2015;167:562-+.
25. Hough R, Archer D, Probert C. A comparison of sample preparation methods for extracting volatile organic compounds (VOCs) from equine faeces using HS-SPME. Metabolomics 2018;14:10.
26. Walsh MC, Kliegman RM. Necrotizing enteroclitis - treatment based on staging criteria. Pediatric Clinics of North America 1986;33:179-201.
27. Reade S, Mayor A, Aggio R, et al. Optimisation of Sample Preparation for Direct SPME-GC-MS Analysis of Murine and Human Faecal Volatile Organic Compounds for Metabolomic Studies. J Anal Bioanal Tech 2014;5.
28. Aggio R, Villas-Boas SG, Ruggiero K. Metab: an R package for high-throughput analysis of metabolomics data generated by GC-MS. Bioinformatics 2011;27:2316-2318.
29. Hughes DM, Komarek A, Czanner G, et al. Dynamic longitudinal discriminant analysis using multiple longitudinal markers of different types. Stat Methods Med Res 2016.
30. Komarek A, Hansen BE, Kuiper EM, et al. Discriminant analysis using a multivariate linear mixed model with a normal mixture in the random effects distribution. Statistics in Medicine 2010;29:3267-3283.
31. Hughes DM, El Saeiti R, Garcia-Finana M. A comparison of group prediction approaches in longitudinal discriminant analysis. Biom J 2017.
32. Vitali B, Ndagijimana M, Maccaferri S, et al. An in vitro evaluation of the effect of probiotics and prebiotics on the metabolic profile of human microbiota. Anaerobe 2012;18:386-391.
33. DeMartino A, Zigler DF, Fukuto JM, et al. Carbon disulfide. Just toxic or also bioregulatory and/or therapeutic? Chemical Society Reviews 2017;46:21-39.
34. Guo SS, Gillingham T, Guo YM, et al. Secretions of Bifidobacterium infantis and Lactobacillus acidophilus Protect Intestinal Epithelial Barrier Function. Journal of Pediatric Gastroenterology and Nutrition 2017;64:404-412.

Legends for Figures

Figure 1: Classification scheme used to identify babies with increased risk of NEC

Figure 2: ROC curves for the best five performing VOCs by AUC (95% CIs) using longitudinal discriminant analysis



127x127mm (600 x 600 DPI)

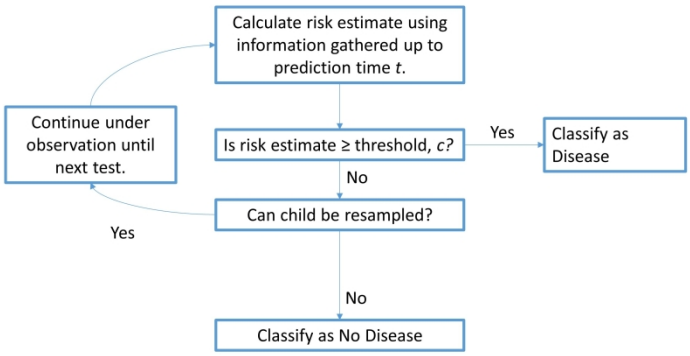


Figure 1: Classification scheme used to identify babies with increased risk of NEC

338x190mm (300 x 300 DPI)

Table S1 Spearman's rank correlation coefficients for compounds in group 4.

	3-methyl-sulfanyl-propanal	2-methyl propanal	Benzene-acetalde-hyde	3-methyl butanal	Benz-aldehyde	2-methyl butanal
Methional	1.00					
2-methylpropanal	0.62	1.000				
Benzeneacetalde-hyde	0.47	0.47	1.000			
3-methylbutanal	0.38	0.49	0.34	1.00		
Benzaldehyde	0.60	0.50	0.45	0.35	1.00	
2-methylbutanal	0.48	0.68	0.38	0.40	0.39	1.00

Table S2 Spearman's rank correlation coefficients for compounds in group 5.

	Butan-2-one	Methanedithione	Heptan-4-one
Butan-2-one	1.00		
Methanedithione	0.39	1.00	
Heptan-4-one	-0.14	-0.21	1.00

Table S3 Spearman's rank correlation coefficients for compounds in group 8.

	2-methylbutanoic acid	3-methylbutanoic acid	Butanoic acid
2-methylbutanoic acid	1.00		
3-methylbutanoic acid	0.68	1.00	
Butanoic acid	0.46	0.33	1.00

Table S4 Samples distribution along time line for analysis showing the number of samples available on each of the 6 days prior to the day of diagnosis

	Number of patients	Day -1	Day -2	Day -3	Day -4	Day -5	Day -6	Total samples
Healthy control	70	34	21	41	17	31	19	163
Confirmed NEC	34	17	9	19	15	16	8	84